Universidade de Aveiro Departamento de Química
Ano 2010

Joana Isabel Monteiro Pinto

Metabonomics of the Blood of Pregnant Women for Diagnosis of Prenatal Disorders

Universidade de Aveiro Departamento de Química
Ano 2010

Joana Isabel Monteiro Pinto

Metabonomics of the Blood of Pregnant Women for Diagnosis of Prenatal Disorders

Dissertation submitted to the University of Aveiro to fulfil the requirements for the degree of Master in the field of Biomolecular Methods, done under the scientific supervision of Doctor Ana Maria Pissarra Coelho Gil, Associated Professor with Aggregation of the Department of Chemistry of the University of Aveiro.

.

Financial support from FCT through project FCT/PTDC/QUI/66523/2006: An NMR-based metabonomics study of health disorders of foetus and mother during pregnancy: towards biochemical characterisation and early diagnostics

Examining Commitee

President

Prof. Dr.ª Maria do Rosário Gonçalves dos Reis Marques Domingues
Auxiliary Professor of the Department of Chemistry of the University of Aveiro

Prof. Dr.ª Isabel Marques Carreira
Associated Professor of Cytogenetics Laboratory and CNC of the Faculty of Medicine of the University of Coimbra

Prof. Dr. Brian James Goodfellow
Auxiliary Professor of the Department of Chemistry of the University of Aveiro

Prof. Dr.ª Ana Maria Pissarra Coelho Gil
Associated Professor with Aggregation of the Department of Chemistry of the University of Aveiro

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Ana Gil for giving me the opportunity to carried out this work and experience first hand the scientific research. I thank for her guidance, help and availability.

I would like to especially acknowledge Dr. António Barros for kindly helping me with multivariate analysis.

My gratitude also goes to Gonçalo Graça for contributing to this work with the NMR experiments and his indispensable help to analyse the resulting data, and to João Rodrigues for helping me with FTIR spectroscopy experiments and data analysis.

I would also like to acknowledge to everyone in the Metabonomics Group of the University of Aveiro who directly or indirectly have contributed to the work presented here.

I wish to knowledge the Fundação para a Ciência e Tecnologia (FCT) for their financial support through the research project FCT/PTDC/QUI/66523/2006: An NMR-based metabonomics study of health disorders of foetus and mother during pregnancy: towards biochemical characterisation and early diagnostics.

I would like to acknowledge the Prenatal Diagnosis Service in Bissaya Barreto Maternity and the Cytogenetics Laboratory of the Faculty of Medicine in Coimbra for the important contribution to the project FCT/PTDC/QUI/66523/2006.

I also acknowledge the Chemistry Department of University of Aveiro for providing the physical conditions for this work.

Finally, I thank to my mother for her patience and to my all family and friends.

Palavras-chave

Metabonómica, Diabetes Gestacional, Malformações Fetais, Plasma Sanguíneo, Ressonância Magnética Nuclear, Infravermelho com Transformadas de Fourier, Análise Multivariada.

Resumo

A aplicação da metabonómica na pesquisa de novos biomarcadores de doenças tem ganho um interesse crescente na investigação e desenvolvimento, tanto ao nível do processamento analítico como do tratamento de dados. Nomeadamente, a análise metabonómica usando espectroscopia de Ressonância Magnética Nuclear (NMR) fornece uma grande quantidade de dados de uma forma rápida e não invasiva sobre a composição de amostras complexas como o plasma sanguíneo. Uma vez que as doenças pré-natais têm um elevado impacto no metabolismo materno e fetal, sendo responsáveis por várias complicações durante e depois da gravidez, esta estratégia foi aplicada ao estudo destas doenças através da análise de sangue de senhoras grávidas (colhido entre 15-24 semanas de gestação), com o objectivo de investigar possíveis metabolitos marcadores ou com poder de previsão para a diabetes gestacional e malformações fetais. Num primeiro passo, foram estudados os perfis metabólicos em RMN dos controlos (n=20) e gravidezes com diagnóstico ou suspeita de malformações fetais (n=11) e pré-diabetes gestacional (com posterior diagnóstico clínico entre 22-34 semanas de gestação). A análise multivariada (análise de componentes principais, PCA; análise discriminante pelo método de mínimos quadrados parcias, (PLS-DA) e duas versões deste último, interval PLS-DA e ortogonal PLS-DA (OPLS-DA)) foram aplicados com o objectivo de pesquisar por correlações de solidez estatística entre a composição do plasma e a ocorrência das doenças em estudo. Os resultados mostraram que as amostras controlo e doença podem ser diferenciadas com base no seu perfil metabólico, nomeadamente mostrando níveis mais elevados de compostos que contêm colina em mulheres que desenvolveram diabetes gestacional mais tarde na gravidez. Adicionalmente, níveis mais elevados de piruvato, manose e compostos que contêm colina, e níveis mais baixos de vários aminoácidos e acetato foram encontrados nas gravidezes afectadas por malformações fetais. Numa segunda etapa do trabalho, as mesmas amostras foram analisadas por espectroscopia de Infravermelho com Transformadas de Fourier (FTIR), um método mais barato e acessível para eventual uso clínico. O perfil dos espectros de FTIR também revelou algumas diferenças entre controlos e doenças, no entanto a sua interpretação específica torna-se difícil devido à grande sobreposição de bandas característica de espectros de infravermelho. Estes resultados mostraram que a análise metabonómica de plasma de mulheres grávidas por RMN e FTIR pode ser uma ferramenta poderosa para obter informação bioquímica sobre a saúde pré-natal e encontrar possíveis novos marcadores com potencial para prever doenças, particularmente no caso do diabetes gestacional.

Keywords Metabonomics, Gestational Diabetes, Fetal malformations, Blood Plasma, Nuclear Magnetic Resonance, Fourier Tansform Infrared, Multivariate Analysis.

Abstract

The use of Metabonomics to search for new disease biomarkers has gained increasing interest in the research community and continuous developments, both at the analytical and data processing levels have boosted this area into new quests in biomarker research. Namely, Nuclear Magnetic Resonance (NMR)-metabonomics provides a large amount of compositional data on complex samples such as blood plasma, in a rapid and non-invasive manner. Since prenatal diseases have a high impact on both maternal and fetal metabolisms, being responsible for a range of complications both during and after pregnancy, this strategy was hereby applied to the study of prenatal diseases, through the analysis of blood (collected at 15-24 gestational weeks), in order to probe for possible marker/predictor metabolites for gestational diabetes and fetal malformations. In the first stage of this work, the plasma metabolic profiles of controls (n=20) and pregnancies affected by diagnosed or suspected fetal malformations (n=11) and pre-gestational diabetes (with *posterior* clinical diagnosis at 22-34 gestational weeks) were evaluated by NMR spectroscopy. Multivariate analysis (principal component analysis, PCA; partial least squares discriminant analysis, PLS-DA and two extended versions of the latter, interval PLS-DA (iPLS-DA) and orthogonal PLS-DA (OPLS-DA) were applied in order to search for consistent statistical correlations between plasma composition and the occurrence of the diseases. It was found that controls and diseased subjects could be differentiated with basis on their plasma profile, namely showing higher levels of choline-containing compounds in pre-gestational diabetic women. In addition, higher contents of pyruvate, mannose and choline-containing compounds and lower contents of several amino acids and acetate were found in pregnancies affected by fetal malformations. In a second stage of the work, the same samples were analysed by Fourier Transform Infrared (FTIR) spectroscopy, a cheaper and more-accessible method, more suited to straightforward clinical use. The FTIR spectral profiles also revealed some differences between controls and diseased subjects, the interpretation of which posing a harder challenge than that of NMR. These results have shown that NMR and FTIR metabonomics of pregnant women blood plasma may be a powerful tool to gain insight into prenatal diseases and find possible new markers with potential predictive value, particularly in the case of gestational diabetes.

Contents

Abbreviations

ATR	Attenuated Total Reflectance
CATS	Chemometrics Analysis Tool System
COSY	Correlation Spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill
DOSY	Diffusion-ordered Spectroscopy
FTIR	Fourier Transform Infrared Spectroscopy
FM	Fetal Malformations
GD	Gestational Diabetes
GC	Gas Chromatography
HDL	High Density Lipoprotein
HMBC	Heteronuclear Multiple Bond Correlation
HMQC	Heteronuclear Multiple Quantum Correlation
HSQC	Heteronuclear Single Quantum Correlation
iPLS-DA	Interval Partial Least Squares-Discriminant Analysis
JRES	J-resolved spectroscopy
LC	Liquid Chromatography
LDL	Low Density Lipoprotein
MS	Mass Spectrometry
MVA	Multivariate Analysis
NMR	Nuclear Magnetic Resonance Spectroscopy
OPLS-DA	Orthogonal Partial Least Squares-Discriminant Analysis
PCA	Principal Component Analysis
PLS-DA	Partial Least Squares-Discriminant Analysis
TOCSY	Total Correlation Spectroscopy
VIP	Variable importance in the projection
VLDL	Very Low Density Lipoprotein
UVA	Univariate Analysis

1.1 Pregnancy and Prenatal Diseases

Pregnancy is an extraordinarily dynamic metabolic state with very important maternal-placental-fetal metabolic interactions. During pregnancy, some metabolic adaptations take place in the mother's metabolism to ensure the continuous supply of nutrients to the fetus, in order to maintain its adequate growth and development.[1]

The placenta is the primary site of nutrient and gas exchange between the mother and fetus. It has three main functions: ensuring fetal metabolism (e. g., synthesis of glycogen), transport of gases and nutrients and endocrine excretion. Only selected nutrients from the maternal blood are able to cross the placenta, glucose being the main carbohydrate transported and a primary source of energy to the fetus. Glucose is the nutrient responsible for maintenance of basal metabolism, energy storage in form of glycogen and in adipose tissues, as well as providing energy for protein synthesis and growth. In addition, the active transfer of amino acids from maternal circulation ensures the structural basis for protein synthesis, essential for fetal growth, and provides the oxidative substrates for energy production, mainly when glucose levels are low.

Vitamins also play an important role for fetal growth and development. Lipids cross the placenta in much lower proportion, however, maternal lipid metabolism undergoes many changes during pregnancy in order to satisfy both maternal and fetal needs. However, not only nutrients and gases have the ability to cross the placenta but also most of drugs, chemicals and infection agents, that can be harmful for fetal health and development.[1-2] In addition, other prenatal complications such as gestational diabetes, preeclampsia, spina bifida, several chromosomapathies, and other abnormalities, may occur during pregnancy with varying impacts on both maternal and fetal metabolisms.

Preeclampsia is one of the most common complications in pregnancy and a major cause of maternal and fetal morbidity and mortality. Although considerable research has been carried out, the cause or causes of preeclampsia remain unclear. Some studies have suggested that the reasons for the development of preeclampsia are increased triglycerides levels, delayed triglycerides clearance, and high blood pressure.[3] Furthermore, preeclamptic pregnancies are usually associated with enhanced hyperlipidemia, which seems to have an impact on fetal lipid profile as evaluated in umbilical cord blood.[4] Preeclampsia is detected by measurement of elevated blood pressure and protein in urine and the only cure known is delivery.[5]

Gestational diabetes (GD) is another very common complication in pregnancy, being associated with many pregnancy complications including preeclampsia. There are some studies that found relations between hyperlipidemia, elevated concentrations of total cholesterol and triglycerides with GD.[6-7] Thus, it seems very important that maternal

blood lipids are regularly evaluated during prenatal care in order to detect abnormal modifications that could be associated to substantial risks for mother and fetus.

Congenital abnormalities can be caused by different factors such as genetic (chromosome abnormalities), environmental (drugs and viruses), and multifactorial inheritance. They are an important cause of prenatal, perinatal, and infant mortality and morbidity.

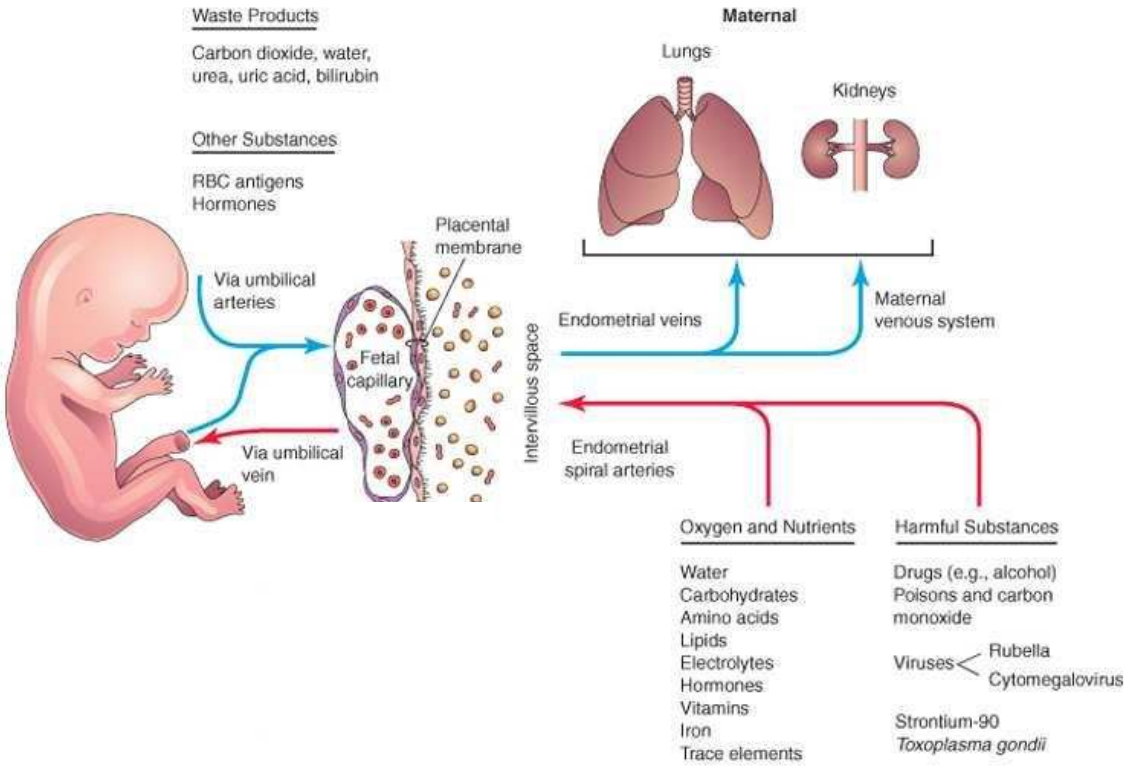


Figure 1.1 Transfer processes taking place between maternal blood, placenta and fetus. The placental membrane consists of extrafetal tissues across which transport of substances between the mother and fetus occurs. [Adapted from reference 2]

Several prenatal tests are used to assess the maternal health during prenatal care. They usually include a complete blood count or haemoglobin and hematrocit test, Rh type and ABO blood group, red cell antibody screening, rubella immunity status, glucose challenge testing, urinalysis, maternal serum alpha-fetoprotein or maternal triple screen, hepatitis B testing, culture for sexually transmitted diseases, surveillance for group B streptococci, and human immunodeficiency virus (HIV).[8] Additionally, many types of non-invasive and invasive techniques are used to determine the healthy or any abnormality of the fetus as can be seen in Table 1.1.[2, 9]

Ultrasonography is the main non-invasive tool for prenatal diagnosis of fetal abnormalities and has the advantages of being low cost, widely available, and free of

known adverse effects. However, in order to detect genetic disorders such as Down syndrome, Edward syndrome, Turner syndrome and rare, inherited metabolic disorders, invasive procedures are usually used when an elevated risk is indicated by a non-invasive screening test.[2]

The most common invasive prenatal diagnostic procedure is amniocentesis. Other invasive procedures used are chorionic villus sampling and percutaneous umbilical cord blood sampling. These very invasive methods have a percentage of risk for mother and fetus because a needle is inserted through the mother's abdominal and uterine walls into the uterine cavity. Another disadvantage is the typically high length of time (weeks) required to the results of amniocentesis and chorionic villus sampling diagnostic procedures to become available.[2, 9]

Table 1.1 Prenatal screening and diagnostic procedures.[9]

	Description	Timing of procedures	Length of time for results	Some disadvantages
Maternal serum marker screening	Screening test for neural tube defects and chromosome anomalies	15-20 weeks	3-7 days	False-positive and false-negative results
Ultrasonography	Screening test for genetic disorders; diagnostic procedure for structural anomalies	4-40 weeks	Immediate	
Chorionic villus sampling	Diagnostic procedure for chromosome anomalies and genetic disorders	10-12 weeks	7-14 days	0.5-2% fetal loss rate; 2% ambiguous results
Amniocentesis	Diagnostic procedure for chromosome anomalies and genetic disorders	15-20 weeks	7-14 days	0.5-1% fetal loss; potential for 2nd trimester termination of pregnancy
Percutaneous umbilical blood sampling	Diagnostic procedure for fetal disorders that require analysis of blood serum levels	19-40 weeks (typically performed at 19-21 weeks)	1-5 days	2-5% fetal loss; potential for 2nd trimester termination of pregnancy

Due the lack of screening methods for preeclampsia, the highly invasive methods for congenital malformations, more research is required with the aim to develop reliable, fast and less invasive diagnostic procedures. It is hoped, therefore, that the study of prenatal diseases with the aim to understand the biochemical metabolism associated may help to advances in prenatal diagnosis.

The work carried out in the scope of this thesis focused the search of metabolic markers of GD, and fetal malformations using NMR metabonomics to study the metabolic profile of maternal blood plasma. A limiting factor in this study was the low number of samples, which unfortunately is a common stumbling block in metabonomics. Hence, only GD and fetal malformations groups out of the groups of disorders interest were considered but there is hope that in future it may also be possible to study other prenatal diseases such as preeclampsia and genetic disorders like Down syndrome. The disorders considered in this study will be discussed in more detail in the next two sections.

1.1.1 Gestational Diabetes

GD is defined as "a carbohydrate intolerance of varying degrees of severity with onset or first recognition during pregnancy". It occurs in 1-14% of all pregnancies, depending on the population studied and the criteria used for diagnosis. GD is a heterogeneous disorder that often occurs in subjects with no history of diabetes, and factors such as age, obesity, and genetic background contribute to the severity of the disease.[6]

GD is usually associated with both insulin resistance and impaired insulin secretion. As pregnancy advances, the increasing tissue resistance to insulin creates a demand for more insulin. In the great majority of women, insulin requirements are readily met and, as a result, the balance between insulin resistance and insulin supply is maintained. Nevertheless, if resistance becomes dominant due to impaired insulin secretion, the development of hyperglycemia occurs. Many studies have documented an increase in preeclampsia, polyhydramnios and operative delivery in GD pregnancies.[6, 10]

Infants of mothers with GD are not at increased risk for congenital anomalies, however, they have an increased risk of fetal overgrowth, macrosomia (excessive birth weight), hyperbilirubinemia, birth trauma or hypoglycaemia.[10]

Women with GD also have a significant risk 20-80% of developing diabetes later in life. For this reason, but also for preventing adverse perinatal outcomes, the diagnosis and timely treatment of GD during pregnancy is very important. GD is usually detected in the second trimester between 24 and 28 weeks of gestation an its diagnosis is similar to non-gestational diabetes, that is performed with a 50g oral glucose load followed by a 1 hour venous plasma glucose levels. If the test gives ≥140 mg/dL of glucose, the patient is scheduled for a 3 hour, 100 g oral glucose tolerance test (OGTT) which carries considerably discomfort for the patient. Two or more abnormal values on the OGTT are considered a diagnosis of GD. Patients with high risk factors, such as history of a prior

macrosomic baby, chronic steroid use or a strong family history of diabetes, may benefit from earlier testing, before or at 20 weeks gestation [10].

However, the diagnostic criteria for GD have been controversial for more than three decades with respect to: its clinical significance and benefit of treatment, as well as the optimal strategies for detection and diagnosis and appropriate treatment goals and methods. For example, many researches have concluded that the diagnostic criteria for GD have limitations because they are based on the prediction of subsequent development of diabetes rather than perinatal outcomes such as cesarean delivery, macrosomia, preeclampsia, among others. The strategy to limit the blood glucose testing to "high risk subjects" such as women with severe obesity or with strong family history of diabetes, have also been object of controversy. The controversial issues will remain until the diagnosis criteria for GD can be formulated based on the specific relationships between maternal glycemia and perinatal outcome.[1, 10]

The management of GD includes dietary therapy, exercise, and medical treatment. In dietary therapy, GD patients are advised to have five meals a day and the daily calorie requirement is calculated according to the standard of 30-35kcal/kg. The proportions of carbohydrate, protein, and fat are 40-50, 25-30, and 25-30%, respectively. If dietary control is carried out strictly, most patients do not need medical treatment like insulin therapy. Exercise may increase the sensitivity of peripheral tissues to insulin, thus being able to help to control glucose level of GD patients.[1]

1.1.2 Fetal Malformations

In the field of embryology, a malformation is defined as a "morphologic defect of an organ, part of an organ, or larger region of the body that results from an intrinsically abnormal developmental process". In some pregnancies, abnormalities of the organ systems occur during fetal development and these comprise neural tube defects, ventricular septal defects, amelia, cleft lip, cleft palate, low-set malformed ears and deafness among others. Some of these structural anomalies occur early in the main embryonic period (3-8 weeks of gestation) when organ systems are formed, but also later during the growth and maturation of the fetus (9-38 weeks of gestation), when organs such as the cerebral cortex, the gastrointestinal tract and the renal glomeruli continue to differentiate and develop.[2]

More than 20% of infant deaths in North America are attributed to birth defects. For example, spina bifida cystic, a severe type of vertebral in which part of the neural tube fails to fuse, are observed in approximately 3% of newborn infants.[2] Congenital anomalies caused by genetic factors also have a high incidence. Chromosomal abnormalities such as Turner syndrome (monosomy, 45 X), Down syndrome (trisomy 21), Edwards syndrome

(trisomy 18), and Patau syndrome (trisomy 13), are present in 6% to 7% of zygotes (single-cell embryos).[2]

From 50% to 60% of spontaneous abortions (most occuring during the first 3 gestational weeks) are due chromosomal abnormalities. There is also a higher incidence of fetuses with neural tube defects, cleft lip and cleft palate.[2] In Table 1.1, general screening and diagnostic procedures to identify congenital abnormalities are described. Ultrasonography usually identifies many structural abnormalities such as neural tube defects but is less accurate for cardiac anomalies. For genetic disorders, for example Down syndrome, screening may be offered in the first trimester (e.g., nuchal translucency scan by ultrasonography and maternal blood tests), or the second trimester (e.g. maternal triple blood screen). Women with high risk are offered a diagnostic test, amniocentesis and chorionic villus sampling.[9, 11]

1.2 General Blood Composition and Relation to Disease

Human blood is the most important transport medium in the body and constitutes about 8% of body's weight. It transports gases, nutrients and wastes, ensuring a balanced distribution of water (homeostasis), and playing a decisive role in body defences against pathogens. Blood interacts with all tissues and, hence, contains a sort of physiological average information about the biochemical status of the body.[12]

The components of blood can be separated into cells and an aqueous medium, the blood plasma, by ultracentrifugation (Figure 1.2). The cellular elements comprise erythrocytes (red blood cells), leukocytes (white blood cells), and thrombocytes (platelets). Blood plasma is composed of water and solutes. The solutes comprise inorganic ions (Na^+, Ca^{2+} ans Cl^-), proteins, lipoproteins, nutrients, metabolites, waste products, and hormones. Plasma proteins are the most abundant plasma solids, being responsible for transport, regulation of water balance, hemostasis and defense against pathogens. Albumin is the protein with highest concentration in blood plasma playing a crucial role in maintaining the blood's colloid osmotic pressure and representing an important amino acid reserve. Lipoprotein complexes are involved in the transport of lipids through the blood stream, from their site of absorption or synthesis (mainly the liver and intestine), to peripheral tissues. The remaining metabolites are in constant flux between blood and tissues. Oxygen transport to maintain the oxidative metabolism of most tissues occurs constantly, as well as carbon dioxide transport from the tissues to lungs. Nutrients absorbed in digestive tract, such as glucose, amino acids, and minerals, circulate to all body tissues, while, waste products in tissues, such as urea and creatinine, circulate through the kidneys and are excreted in urine. Hormones are also transported in plasma from endocrine glands to their target organs.[13-14]

Since blood is an important vehicle to establish the diagnosis of a patient providing the molecular information on biochemical status, laboratory analysis of blood samples are part of the diagnostic process of several diseases such as diabetes, liver diseases, and cancer, among others.[15] However, additional studies of the metabolic profile of biological samples have proved to be of further use to study several diseases.[16-19] This strategy, known as metabonomics, is very attractive because many metabolites and their biochemical processes can be investigated rather than only particular metabolites. The information thus obtained can be used not only for disease diagnosis but also for prognosis, treatment follow-up and monitoring the efficacy and safety of drug treatment.[20]

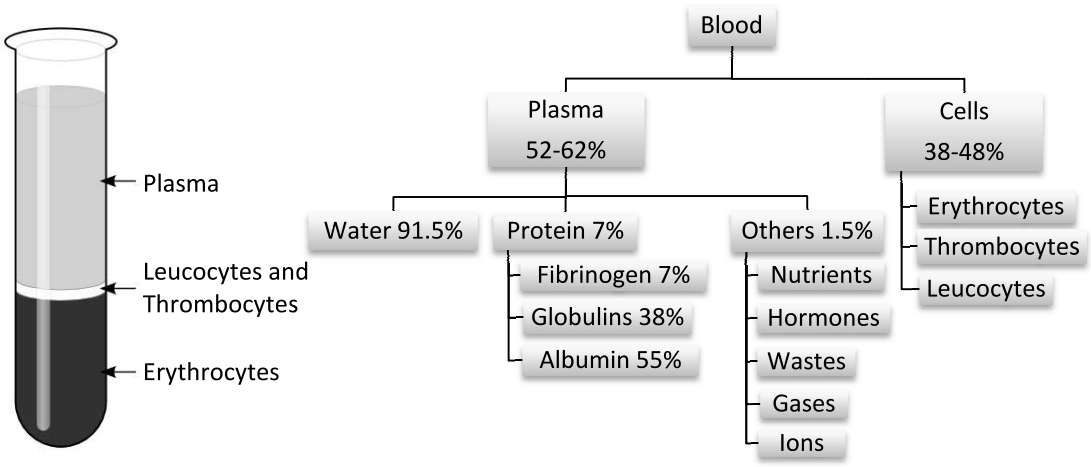


Figure 1.2 Centrifuge tube after blood centrifugation showing three layers, and main components of blood.[13]

Over the years, many different diseases have been associated with a disturbance in lipoprotein balance and composition. Besides atherosclerosis and coronary heart disease (CHD)[21-23], dysregulations in lipoprotein metabolism are critical to development of cancer[24-26], diabetes[27-30] and different types of hyperlipidemia[31]. Prenatal diseases such as preeclampsia and GD also have been associated with changes in maternal lipoprotein metabolism.[3-4, 6-7, 32-33] For these reasons, the study of lipoprotein particles was carried out in the scope of this thesis and some useful general information about their function and structure is presented in the next paragraphs.

Lipoproteins are usually classified on the basis of their density on isolation by ultracentrifugation. The main fractions are chylomicrons (CM), very low-density (VLDL), intermediate-density (IDL), low-density (LDL), and high-density (HLDL) lipoproteins.

The general structure of lipoproteins is formed by an envelope and the core (Figure 1.3). The envelope comprises mainly apolipoproteins and amphipathic lipids (mostly phospholipids and unesterified cholesterol) with their polar groups oriented to the aqueous surface, while the apolar groups interacts with the hydrophobic core, formed by triacylglicerols, cholesteryl esters, and small amounts of unesterified cholesterol and other dissolved lipids.[34]

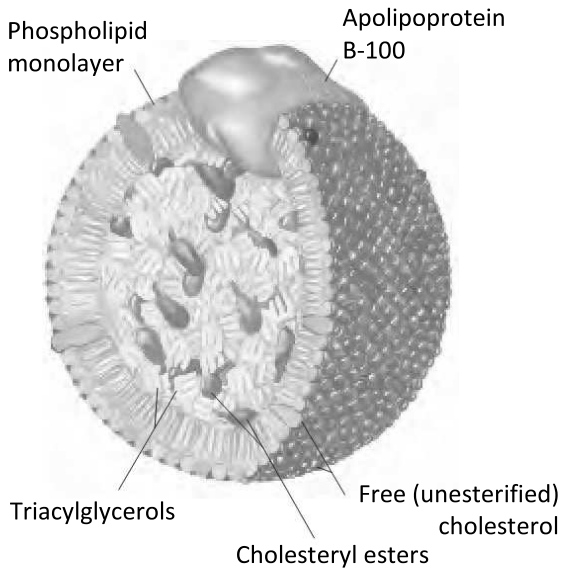


Figure 1.3 Structure of a low-density lipoprotein (LDL). [Reprinted from reference 34]

As can be seen in Table 1.2, the lipoprotein fractions have similar chemical composition but differ in relative lipid and protein composition and, hence, different combinations of lipids and proteins produce particles with different densities, size and functions. The lipid content is usually inversely correlated with the density of the lipoproteins.

The differences in contents between lipoprotein fractions are related to their functional properties. Chylomicrons transport dietary triacylglycerols from intestines to other tissues such as adipose, heart, skeletal, and muscle. VLDL export endogenous triacylglycerols from the liver. LDL are the major carriers of cholesterol in blood and deliver cholesteryl esters to peripheral tissues and liver. Finally, HDL remove excess of cholesterol from cells and transports it to liver and steroidogenic tissues for metabolism and secretion.[34]

Table 1.2 Composition of major classes of human plasma lipoproteins.[12]

			Composition (wt %)*				
			Surface Components			**Core Lipids**	
Lipoprotein Class	*Diameter (nm)*	*Density (g/mL)*	*Protein*	*Phospholipids*	*Cholesterol*	*Cholesteryl esters*	*Triacylgly-cerols*
Chylomi-crons	75 -1200	0.93	2	7	2	3	85
VLDL	30 - 80	0.93 - 1.006	8	18	7	12	55
IDL	25 - 35	1.006 – 1.019	19	19	9	29	23
LDL	18 - 25	1.019 – 1.063	22	22	8	42	6
HDL2	9 - 12	1.063 – 1.125	40	33	5	17	5
HDL3	5 - 9	1.125 – 1.21	45	35	4	13	3

Analytical methods such as ultracentrifugation, electrophoresis, chromatography, precipitation, and enzymatic assays, are usually applied for lipoprotein analysis. However, they are time and labour consuming, some of them requiring large amounts of plasma and sometimes introducing artefacts by modifying particles and their constituents.[35]

NMR spectroscopy has proven to be useful for characterization of the main lipoprotein particles using liquid chromatography directly coupled to NMR (LC-NMR)[36] and mathematical separation applied to 1D ^{1}H NMR spectra.[37-41] These approaches will be discussed in more detail in the next subchapters.

1.3 The Metabonomics Approach

1.3.1 General Strategy

Metabonomics is defined as "the quantitative measurement of the multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification".[42] This concept emerged from the use of ^{1}H NMR spectroscopy to study the metabolic composition of biofluids, cells and tissues in the mid-1980s.[43-45] The metabonomics approach deals with detecting, identifying, quantitating and cataloguing the history of metabolic changes in complex biological samples creating comprehensive global metabolic profiles.[42] In order to achieve these goals, it combines the use of analytical technologies and multivariate statistical methods. Several analytical methods have been used to generate metabonomics data sets, such as ultraviolet, infrared rarely used, but the main ones are NMR spectroscopy, liquid chromatography or gas chromatography couplet with mass spectrometry (LC-MS and GC-MS), for generating metabolic profiles. Multivariate analysis (MVA) usually involves the application of unsupervised methods such as principal component analysis (PCA), hierarchical cluster analysis, and nonlinear mapping, and supervised methods such as partial least squares (PLS), soft independent modelling of class analogy, and neural networks, with the aim to interpret profiles of metabolism.[20]

The most successful approaches have been NMR spectroscopy and MS (Figure 1.4), as these methods can capture information on hundreds or even thousands of metabolites in a sample in a single analytical run. NMR spectroscopy provides detailed information on molecular structure of compounds as well as information on absolute or relative concentrations. It has the advantage of being non-destructive, requires little sample preparation, unbiased, robust, reproducible, and offers high sample throughput efficacy (10-15 minutes per sample with a conventional detection probe). However, NMR is less sensitive than MS, but useful data can still be generated from small amounts of samples. MS also provides both metabolic identification and quantitation, however, it requires well adapted sample preparation and, often a prior separation of the complex mixture sample using chromatographic methods is required.[20, 46] MVA is the next step used with the aim to encode the complex metabolic profiles for the identification of biomarkers, and then elaborate of biological outcome.[46]

Fourier transform infrared (FTIR) spectroscopy is a technique rarely used in metabonomics. Compared to NMR, it is cheaper, simpler and a more sensitive way of obtaining molecular information on complex samples. Like NMR, there is no need to preselect analyte classes prior to analysis and can also be used in tandem with MVA. Thus, any differences in spectra due to a perturbation can be interpreted crudely in terms

of the functional groups of the substances involved. However, the main limitation of FTIR is the low level of detailed molecular assignment that can be achieved.[20, 47]

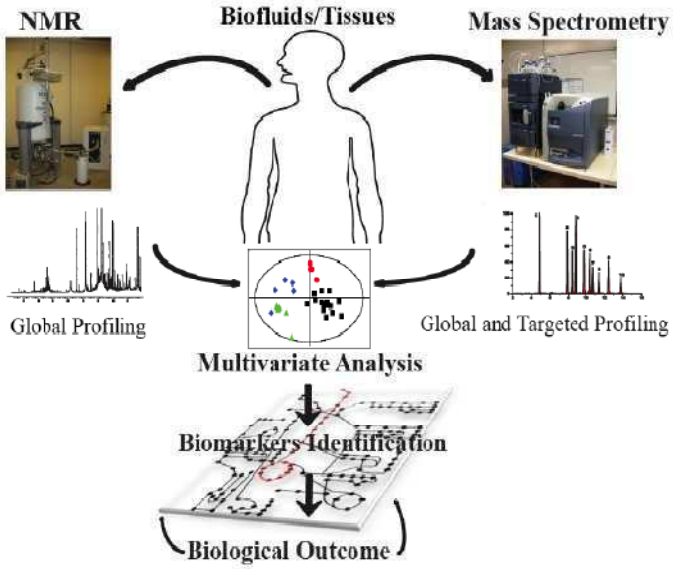


Figure 1.4 Illustrative scheme of a typical NMR (Bruker 600 MHz) and MS (Waters Acquity Ultra Performance Liquid Chromatograph) based metabonomics analytical platform. [Extracted from reference 46]

Currently, many types of biofluids, tissues and cells extracts are studied with the aim to investigate their composition and biochemical information to provide insight into the status of a living organism which can lead to ways of predicting disease evolution.[46] The most used biofluids in metabonomics studies of biomedical relevance are urine and blood plasma. These can be easily obtained for disease diagnosis and can provide an integrated view of the whole systems biology. However, often the limiting factor found in biomedical studies of diseases is the low number of samples available.[20]

The more extensive application of metabonomics has been in drug toxicology, studying the effects of drugs on whole organisms by the measurement of alterations in the metabolic profile over time, in response to stressor or intervention. In this respect, metabolic nuances and biomarkers, can be identified by characteristic changes in the pattern of concentrations of endogenous metabolites that relate to the site and mechanisms of toxicity.[48] A ^{1}H NMR metabonomics biomarker is defined as an area(s) which is (are) consistently changed in the ^{1}H NMR spectral region by a given factor of interest (disease, toxicology).[49] NMR-based metabolic profiling has also been widely used in human disease diagnosis to derive new biochemically based assays and indentify combination biomarkers for diseases. For instance, there is a great need for sensitive and objective testing methods for cancer within the asymptomatic population, hence much research has been carried out in related areas such as breast[50-51], liver[52] and lung[53-54]

cancers. Coronary heart disease has also been investigated due to its high incidence of mortality in developed countries.[55-56]

The study of prenatal disorders by metabonomics has been addressed before, using maternal plasma samples[57-58] and human amniotic fluid.[59-61] For example, significantly higher concentrations of histidine, tyrosine and phenylalanine were found in preeclamptic pregnancies compared to normal pregnancies.[58] These findings have already been applied as diagnostic markers by a company called Metabio ltd from the University of Leeds (established in 2008).

1.3.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR is a spectroscopic technique based on the magnetic properties of the atomic nuclei when placed in a strong magnetic field, and provides information on the molecular structure of organic molecules and biomolecules in solution.[62] NMR is a more rapid and non-invasive technique. For these reasons, NMR spectroscopy has been widely used in structural chemistry and, more recently for metabolic profiling, not only to access the status of living organisms[54, 63] but also for detailed characterization of foods such as fruit juices and beer.[64-65]

1.3.2.1 Principles of NMR Spectroscopy

The nuclear spin generates a magnetic dipole along the axis of rotation, this spin having a magnetic moment, μ, associated as well as an angular moment, P. Both, angular moment P and magnetic moment μ are related to each other by $\mu = \gamma P$, where γ is a characteristic of the nucleus called gyromagnetic ratio. The angular moment of spin depends on the nuclear spin quantum number, I, which is greater than zero for nuclei presenting spin being therefore "NMR visible". For ^{1}H the spin quantum number is 1/2.

When samples are exposed to an external static and homogeneous magnetic field referred to as B_0, the magnetic moments in the sample align along B_0 according to a Boltzmann distribution. In the case of proton, there are two possible spin orientations or two different energy states. It can only align parallel (up) and anti-parallel (down) with respect to the external field. These two possible orientations (Figure 1.5) whose the corresponding magnetic quantum numbers of m_1=+1/2 and m_2=-1/2, usually denominated by α and β respectively.[66]

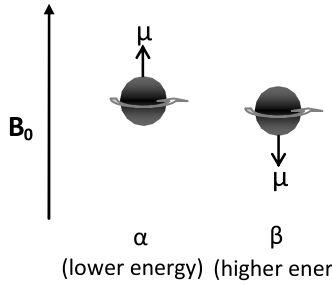


Figure 1.5 Two possible orientations of a spinning proton in the presence of an external magnetic field B_0. Adapted from [67]

The difference of potential energy between these two spin states is given by $\Delta E = E_\alpha - E_\beta = \frac{\gamma h B_0}{2\pi}$. If the magnetic moments of all nuclei present in the sample are added vectorially, the difference of populations between α and β states for a nucleus with spin 1/2 leads to a net vector in the +z direction. The sum of all individual spins is called total magnetization, M_0.[67-68]

In order to obtain a NMR signal, M_0 is flipped orthogonally to B_0 by use of a radio-frequency field B_1 applied for a defined time period (B_1-pulse). For classical physics, M_0 aligns along the *x*-direction and will precess with a resonance frequency given by $w_0 = \frac{\gamma_X}{2\pi} B_0$. Upon energy absorption, there is an exponential decrease of the magnetization which induces a current in the receiver coil that also decays exponentially. This process is recorded as a signal called Free Induction Decay (FID). To convert this accumulated FID signal into a real spectrum, a mathematical process called Fourier transformation is applied. Then, FID signals, which are time-dependent functions stored in a computer memory, are converted to frequency domain. So, we obtain a NMR spectrum, the plotting of signal intensity versus frequency.[20, 66] Nowadays, an NMR spectrometer uses a superconducting magnet which has high field strengths of 3.5-18.8 tesla (T) (150-800 MHz for protons) and provides high sensitivity and stability.[68]

The ^{1}H NMR spectra show many NMR absorptions depending on the electronic environment of the nuclei. Chemical shifts are measured with reference to the absorptions of protons of reference compounds like tetramethylsilane (TMS), which is set to zero. The range in which most of the ^{1}H NMR absorptions occur is relatively narrow, δ 0-14 downfield from the absorption of TMS. The general ranges of absorptions of different functional group can be seen in Figure 1.6.[66]

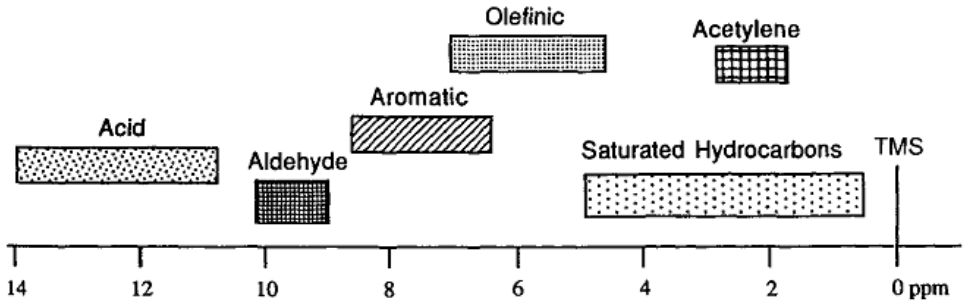


Figure 1.6 General absorption regions of different functional groups. [Reprinted from reference [66]]

1.3.2.2 1D and 2D ^{1}H NMR Spectroscopy in Metabonomics

NMR spectroscopy has the extraordinary ability to provide detailed information on molecular structure of both pure compounds and complex mixtures such as foods or biofluids. However, the assignment of ^{1}H NMR spectra of a mixture can be a complex procedure, and usually requires several 1D and 2D NMR methods, as well as, comparison with spectra of standard compounds.

1D NMR experiments are largely used in metabonomics due the rapid acquisition times (few minutes per sample for blood plasma), making it suitable in studies with large sample arrays. Samples like biofluids have a high content of water and, hence, the water protons signal dominates the spectrum obscuring a large section. Thus, the water signal has to be suitable suppressed which is commonly achieved by water suppression pulse sequences such as a simple presaturation, NOESYPRESAT or Watergate sequences.[69] Usually, three types of 1D experiments are acquired in metabonomics studies of complex samples: (i) a 1D pulse-and-acquire experiment, in which large and small molecules contribute to the spectrum with intensity proportional to their concentration; (ii) a transverse relaxation-edited or Carr-Purcell-Meiboom-Gill (CPMG) experiment, in which signals from protons of macromolecules are suppressed resulting in a spectrum which is dominated by the smaller molecules; and (iii) a diffusion-edited experiment based on the difference in the diffusion coefficients between macromolecules and low molecular metabolites, an approach in which only macromolecules contribute to the spectrum.[69]

An example of a standard 1D and CPMG ^{1}H NMR spectrum of human blood plasma is given in Figure 1.7. As can be seen in Figure 1.7(a), the standard 1D ^{1}H NMR spectrum of plasma is very complex because the resonances of low molecular weight metabolites, proteins, lipids and lipoproteins are heavily overlapped.

Macromolecules show broad signals in standard 1D ^{1}H NMR spectrum due to their short spin-spin (transverse) relaxation times (T_2). T_2 relaxation times are inversely proportional to the width of the NMR signal, hence, short T_2 values originate broad signals in the ^{1}H NMR spectrum. On the other hand, small molecules are usually

characterized by long T_2 value and, hence, they appear as narrow signals in the spectrum.[62] The CPMG experiment uses the differences in T_2 to attenuate the signals from macromolecules, and the resulting spectrum is dominated by small molecules (Figure 1.7b and Figure 1.7c).

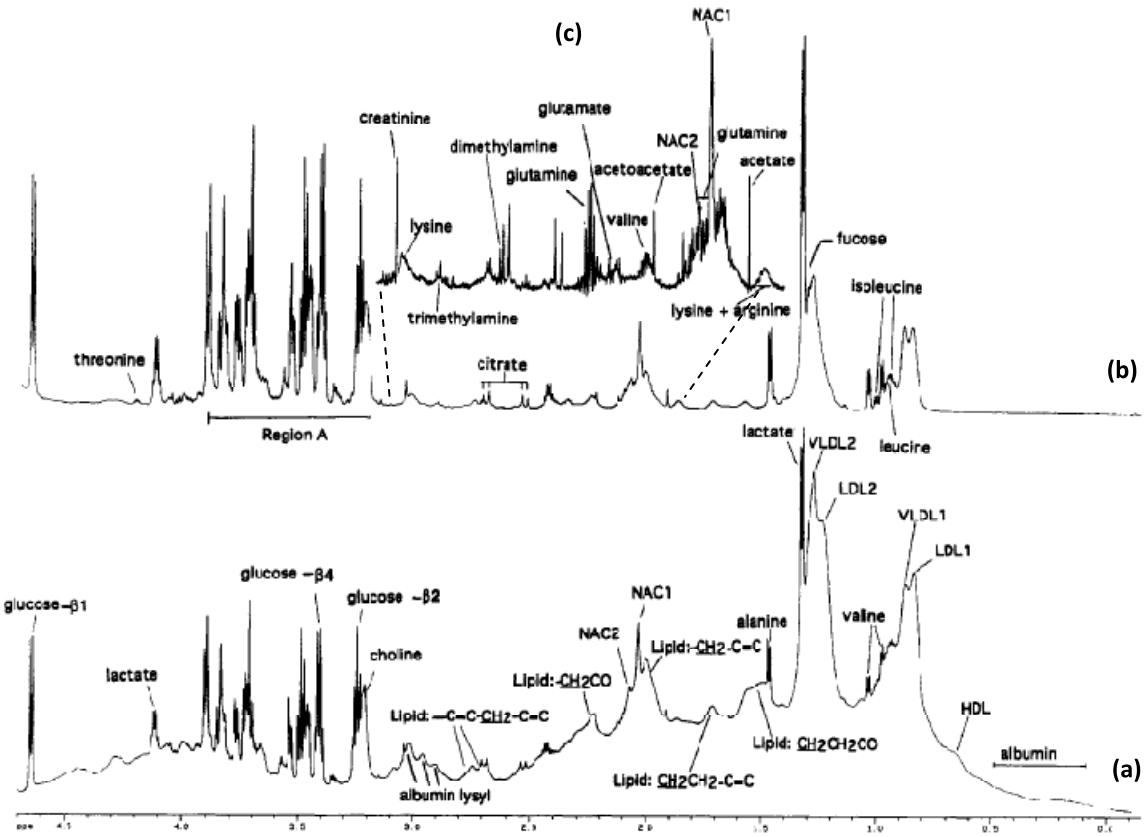


Figure 1.7 750 MHz 1H NMR spectra (δ -0.2 t0 4.7) of human blood plasma at 292 K: (a) standard 1D ^{1}H NMR spectrum, (b) CPMG spin-echo spectrum, and (c) expansion of resolution-enhanced CPMG spin-echo spectrum (δ 1.8 to 3.1). Abbreviations: LDL1 and VLDL1 refer to the terminal CH3 groups of fatty acids in LDL and VLDL, respectively; HDL refers to the C18 signal from cholesterol in HDL; NAC1 and NAC2 refer to the composite signal acetyl signals from α_1-acid glycoprotein. The region in spectrum (a) labelled "albumin" indicates the position of poorly resolved signals from ring current-shifted protons from amino acids in albumin. The area labeled as "Region A" in spectrum B contains signals from glycerol, glucose and amino acid CH protons. [Adapted from reference [70]]

The use of 2D NMR methods is very important to provide additional information to solve overlap problems and allow identification of metabolites that otherwise remain undetected.

2D NMR methods are based on dipolar (through space) or scalar (through bond) couplings between magnetic nuclei. The methods more commonly used for assignment and spectral simplification purposes are J-resolved spectroscopy (JRES), Correlation

Spectroscopy (COSY), Total Correlation Spectroscopy (TOCSY), and inverse-detected heteronuclear correlation methods such as Heteronuclear Multiple Quantum Correlation (HMQC), Heteronuclear Multiple Bond Correlation (HMBC), and Heteronuclear Single Quantum Correlation (HSQC).[20]

The JRES experiment is the simplest 2D NMR experiment and allows spectral simplification based on spin-echo pulse sequence. It separates J-splitting and chemical shifts on to two orthogonal axes: the J-multiplets are displayed along the F1 dimension, while the chemical shifts are shown in F2 dimension. This experiment minimizes overlap because contributions of substances with short ^{1}H T_2 values are suppressed, and allows better determination of coupling constants.

The 2D ^{1}H-^{1}H COSY spectra reveal the network of spin-spin couplings (over 3-5 bonds) in each molecule and can provide simultaneously the chemical shift and scalar coupling by detecting the off-diagonal cross peaks, which are symmetric with respect to the diagonal. The TOCSY spectra are more information-rich than COSY since they detect multiple relayed connectivities, i. e., allow detection of long range interactions (>3 bonds), that are usually too weak in COSY spectra. Additionally, TOCSY spectra have higher sensitivity for large molecules and absorption lineshapes for both diagonal and cross-peaks. However, these spectra can show serious signal overlap on cross-peaks.[20]

Inverse-detected heteronuclear correlation methods are widely used in metabonomics because the chemical shift of the carbons is correlated with the chemical shift of J-coupled protons. Nevertheless, the application in metabonomics is limited due the long acquisition times. HSQC is an example of an experiment that combines ^{1}H and ^{13}C spectroscopy. The chemical shift of ^{13}C is correlated via one-bond coupling constant $^{1}J_{CH}$ to the ^{1}H chemical shift of the directly bond proton. HMQC provides information identical to that from HSQC, while HSQC uses a single quantum transfer, HMQC uses a multiple quantum transfer mechanism.[20] An example of an HMQC spectrum of human blood plasma is shown in Figure 1.8, it allows the assignment of several metabolites in δ 0.5-3.3 region such as cholesterol of lipoprotein particles, and amino acids (e.g. glutamate, lysine) that were not detected in 1D ^{1}H NMR spectra.[70] Finally, in an HMBC experiment, the ^{13}C nuclei with a small multiple bond coupling $^{n}J_{CH}$ are correlated to ^{1}H. For each proton several cross signals related to coupled ^{13}C nuclei appear in the spectrum.[20]

Other 2D experiment useful for assignment of complex samples is the diffusion ordered spectroscopy (DOSY). This experiment allows the discrimination of different components by their diffusion coefficient which is a consequence of their size. DOSY is based on the measurement of a series of spin echo spectra with different field gradient strengths, followed by the analysis of the signals decays to extract a set of diffusion coefficients with which to obtain the diffusion domain of a DOSY spectrum.[71]

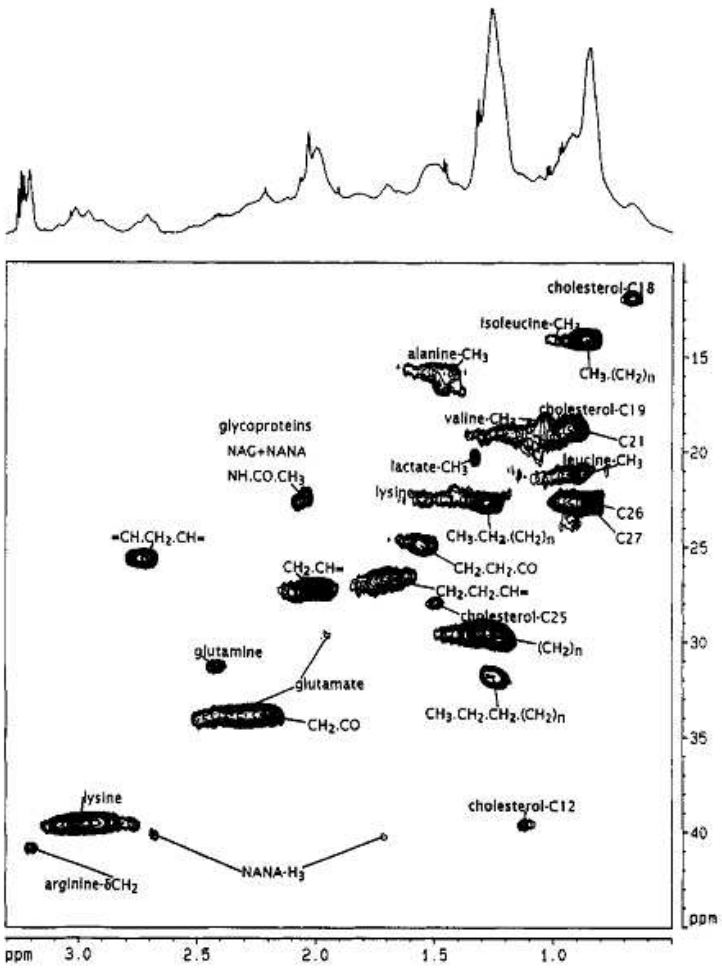


Figure 1.8 750 1H-13C HMQC spectrum of human blood plasma measured at 313 K, showing regions ^{1}H δ 0.5-3.3, ^{13}C δ 10-50. [Reprinted from [70]]

The figures shown above were extracted from the work published by Nicholson et al. in 1995 which provides the most detailed assignment of the ^{1}H NMR spectra of human blood plasma, a list of the assignments and their chemical shifts can be found in Annex 1.[70]

1.3.3 Fourier Transform Infrared (FTIR) Spectroscopy

1.3.3.1 Principles of FTIR Spectroscopy

FTIR spectroscopy involves the interaction of infrared electromagnetic radiation with the test sample. As in other types of spectroscopy, molecules are excited to a higher energy state and absorb selected frequencies (energies), generally correspond to energy changes on the wavenumbers range of 13 000 to 10 cm^{-1}. In most covalent molecules, the radiation in this energy range corresponds to the range encompassing the stretching and

bending vibrational frequencies of the molecule. The condition for a molecule to show infrared absorptions is to possess an electric dipole moment that changes during the vibrations. A bond must present an electrical dipole that is changing at the same frequency as the incoming radiation in order for energy to be transferred. An example of an "infrared-active" molecule is a heteronuclear diatomic molecule such as CO and NO.

The infrared spectrum can be divided into three main regions: the far-infrared (<400 cm^{-1}), the mid-infrared (4000-400 cm^{-1}) and the near-infrared (13 000-4000 cm^{-1}). Every organic molecule exhibits absorption spectra in mid-infrared region (4000 to 500 cm^{-1}). The absorptions of each type of bond are found only in certain small portions of the vibration infrared region. The mid-infrared spectrum can be approximately divided into four regions as follows: the X-H stretching regions (4000-2500 cm^{-1}), the triple-bond region (2500-2000 cm^{-1}), the double-bond region (2000-1500 cm^{-1}) and the fingerprint region (1500-600 cm^{-1}). The Figure 1.9 shows schematically how each type of bond is spread out over the mid-infrared region.[72-73]

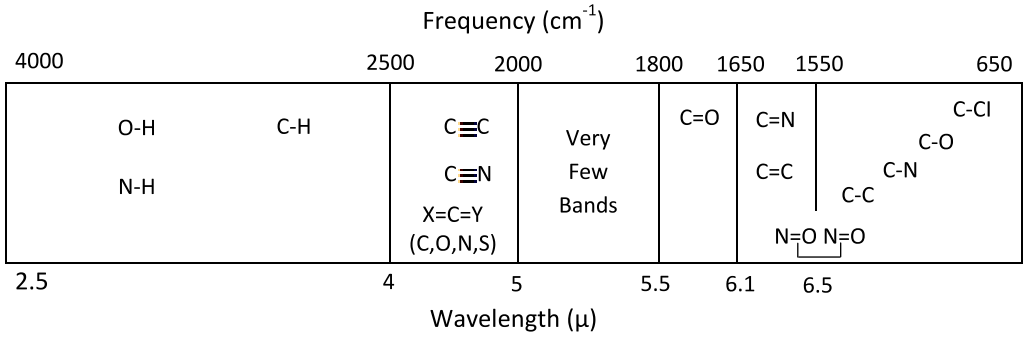


Figure 1.9 The approximate regions where various common types of bonds absorb (stretching vibrations only). [Reprinted from reference [72]]

The basic components of an FTIR spectrometer are shown schematically in Figure 1.10. In the interferometer, the source energy passes through a beam splitter, a mirror placed at a 45º angle to the incoming radiation, which allows the incoming radiation to pass through but separates it into two perpendicular beams, one undefiled, the other oriented at a 90º angle. One beam, the oriented at 90º, goes to a stationary or "fixed" mirror and is returned to the beam splitter. The undeflected beam goes to moving mirror and is also returned to the beam splitter. The motion of the mirror causes the pathlength that the second beam traverses to vary. When the two beams meet at the beam splitter, they recombine, but the pathlength differences (differing wavelength content) of the two beams cause both constructive and destructive interferences. The combined beam containing these interference patterns is the interferogram. Then, the interferogram generated is oriented towards the sample by the beam splitter. At the same time as it passes through the sample, the sample absorbs all of the wavelengths (frequencies) that

are found in its infrared spectrum. The modified interferogram signal that reaches the detector contains information about amount of energy that was absorbed at every wavelength (frequency). The computer compares the modified interferogram to a reference laser beam to have a standard of comparison. Now, the Fourier transform must be applied to extract the individual frequencies that were absorbed, and to reconstruct and plot what we recognize as a typical infrared spectrum.[72]

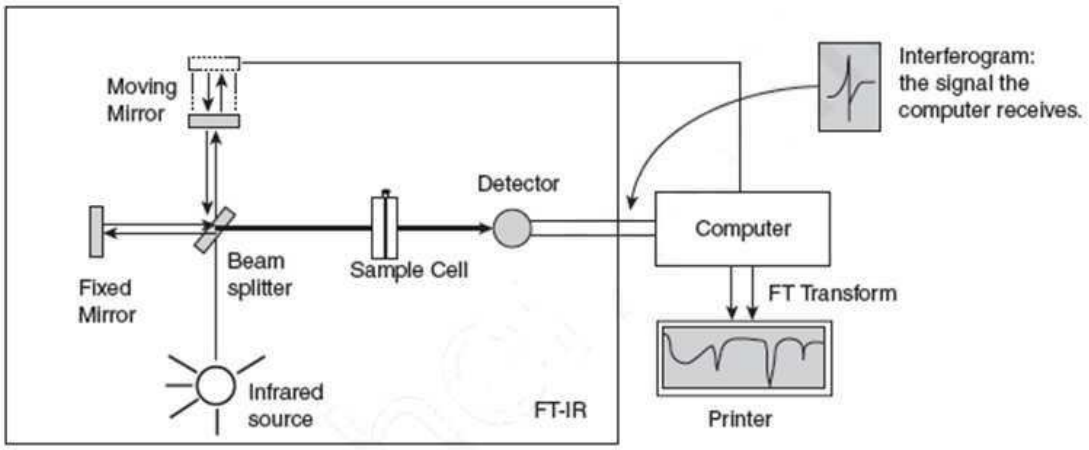


Figure 1.10 Schematic diagram of Fourier transform infrared spectrometer. [Reprinted from reference 72]

Attenuated total reflection (ATR) is a technique used in infrared spectroscopy which utilizes the phenomenon of total internal reflection. ATR consists of reflecting the incidence radiation inside a high refractive index crystal where the sample is placed, and as part of the radiation is absorbed by sample, the radiation loses energy. The resultant attenuated radiation is measured and plotted as a function of wavelength giving rise to the absorption spectral characteristics of the sample. A great advantage of this technique is the minimal or nonexistent sample preparation required and, hence, no changes in morphology are seen.[73]

1.3.3.2 FTIR Spectroscopy in Metabonomics

In spite of the limited used of FTIR spectroscopy in metabonomics studies, its use in studies of complex samples is mainly applied in foods but there are very few successful applications in biofluids. Biofluids such as blood serum and plasma have been investigated for monitoring and identification of metabolic profile in certain types of diseases.[74-76] For example, FTIR spectroscopy and cluster analysis was applied to the analysis of human plasma samples in order to detect spectral parameters, which might serve as biomarkers for monitoring and identification of leukemia patients. The results presented in this study showed several spectral peaks attributed to carbohydrates (1056

cm^{-1}), amide III (1270 cm^{-1}) and amino acids (1592 cm^{-1}), which might be useful biomarkers for detecting such leukemia patients.[75]

There are defined absorption regions in the FTIR spectrum of blood plasma known to be characteristic of some metabolites as in the case of lipids, proteins and saccharides. Figure 1.11 shows a FTIR spectrum of a dried blood plasma sample of an adult man. Due the high water content (91.5%), plasma samples have to be dried before the FTIR spectra acquisition in order to detect the small metabolites. In Table 1.3 is present the assignment of the FTIR spectra of blood plasma.[74]

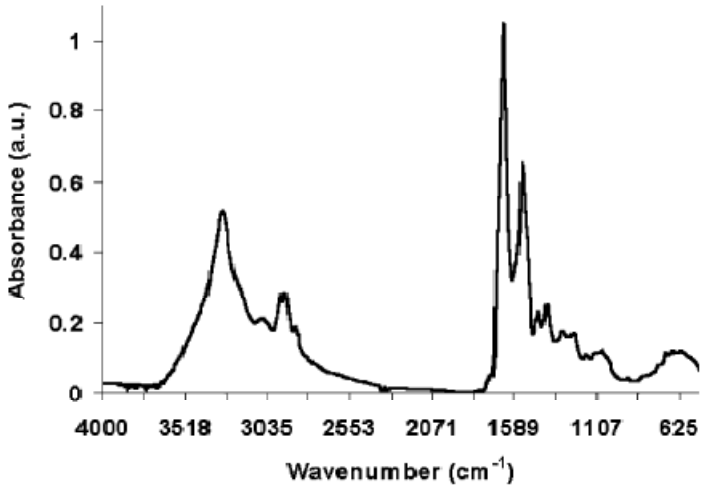


Figure 1.11 FTIR spectrum of an adult man blood plasma obtained from a dried sample. [Reprinted from reference 74]

Table 1.3 Major assignments of absorption bands observed in FTIR plasma spectra.[74]

Bands (cm^{-1})	**Major assignments for plasma contents**
3020-3000	ν(CH) vibration: unsaturated fatty acids, cholesterol esters
2990-2950	$\nu_{as}(CH_3)$ vibration: cholesterol esters, triglycerides
2950-2880	$\nu_{as}(CH_2)$ vibration: long chain fatty acids, phospholipids
2880–2860	$\nu_s(CH_3)$ vibration: cholesterol esters, triglycerides, glycerol
2870–2830	$\nu_s(CH_2)$: long chain fatty acids, phospholipids
2996–2819	$\nu_{as}(CH_3)$, $\nu_s(CH_3)$, $\nu_{as}(CH_2)$, $\nu_s(CH_2)$: fatty acids, phospholipids, triglycerides
1739-1732	νC=O vibration: lipids, cholesterol esters, triglycerides
1720-1600	νC=O vibration: (amide I) β-sheet: proteins, turns, coils
1630–1560	νC=O vibration: amino acids
1600-1480	δN-H vibration: (amide II) α-helix from proteins
1480-1430	$\delta_{as}(CH_3)$, $\delta_{as}(CH_2)$, $\delta_s(CH_3)$, $\delta_s(CH_2)$: fatty acids, phospholipids, triglycerides
1430-1360	$\nu(COO^-)$ vibrations: amino acids
1300-900	N(C-O) vibrations: saccharides, glucose, lactate, glycerol

ν: stretching vibrations, δ: bending (scissoring) vibrations, s: symmetric, as: asymmetric

1.3.4 Multivariate Analysis (MVA): Principles and Tools

1.3.4.1 Outline of MVA of Spectral Data

As mentioned before, MVA applies unsupervised and supervised methods in order to interpret the high spectral complexity obtained in a set of experimental measurements. Unsupervised multivariate techniques like PCA are used to reduce data complexity in a rational way and also produce display plots which can be interpreted by the human eye. Supervised methods such as PLS and orthogonal PLS (OPLS) use the class information given for a training set of sample data to optimize the separation between two or more sample classes.[77] Furthermore, receiver operating characteristics (ROC) graph is a useful technique for visualizing, organizing and selecting classification models based on their performance.[78]

The NMR or FTIR spectra provide an n-dimensional metabolic fingerprint with the coordinate along each dimension being the concentrations of individual measurable compounds or the NMR or FTIR spectral intensity at each data point or collection of data points.[79] Usually, the spectral data is prepared or preprocessed for multivariate modelling. The NMR spectra are frequently divided into regions along the chemical shift axis whose areas are summed to provide an integral so that the intensities of peaks in such defined spectral regions are extracted. This process is called binning and results in a matrix consisting of rows that reflect observations/samples and columns that represent variables. Then, normalization is often applied to the rows which adjusts spectral intensities so that concentration differences between samples are accounted for such that the samples are more directly and reliably comparable. One of the main normalization method applied is normalization to total area or constant sum sets the total spectral area of each spectrum to unity; as a result, the intensities of all data points are expressed relative to this. The final preprocessing step is scaling which is a column operation generally applied to NMR spectral data prior to chemometric modelling to the aim to reduce the noise in data and consequently improve model interpretability, for example, each column in a matrix can be set to have unit variance or a mean of zero.[80]

1.3.4.2 Principal Component Analysis (PCA)

PCA usually gives reliable results though even the number of variables (number of spectral points) exceeds the number of observations (samples) and, also, its high tolerance for collinearity among the variables, which is frequent for NMR data.[81] A two-dimensional data table, called ***X*** matrix, can be constructed containing N observations in a K-dimensional space. The variation in a data table is approximated by a low dimensional model plane that approximates all rows (e.g. observations) in ***X***, and is represented by

points in Figure 1.12. The variation in data matrix ***X*** is described by a scores matrix, *T*, a loadings matrix *P*, and a residuals matrix (E), such that $\boldsymbol{X} = TP^t + E = t_1{p_1}^t + t_2{p_2}^t + E$. The first PCA component ($t_1{p_1}^t$) contains the largest variation in the points. The second PCA component ($t_2{p_2}^t$) is orthogonal to the first and contains the next largest variation and so on. The scores plot (t_1 and t_2), in Figure 1.12, represent the relation among the observations in the model plane and provide an overview of all observations in the data table. Hence, groupings, trends, and outliers are made visible. The loading vectors (p_1, p_2) define the relation among the measured variables that is the columns in X matrix. The loading plot, in Figure 1.12, describes the influence of the ***X***-variables in the model plane. Then, a central feature is that directions in the score plot correspond to directions in the loading plot, and vice-versa. Finally, the residuals matrix (E) is the part of ***X*** that is not explained by the model and represents the distance between each point in K-space and its projection on the plane.[82]

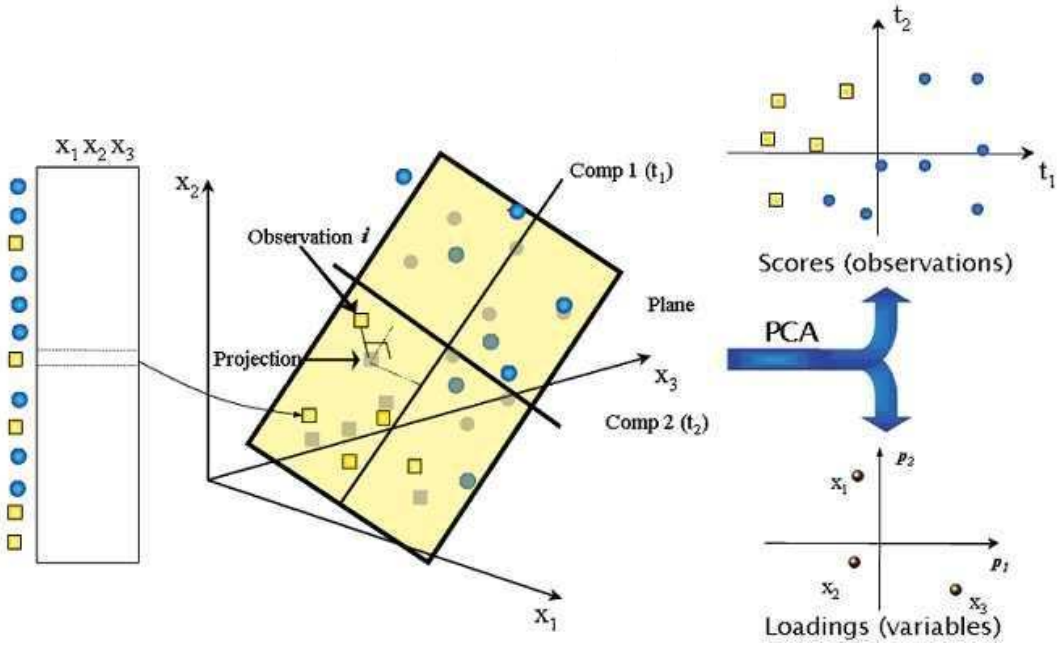


Figure 1.12 Illustration of PCA model. [Reprinted from[82]]

1.3.4.3 Partial Least Squares (PLS) and Orthogonal Partial Least Squares (OPLS)

Multivariate regression methods like PLS have large popularity in a wide range of fields including in metabonomics. While PCA is a method that only considers the structure of one matrix ***X***, PLS constructs a set of orthogonal latent vectors from a matrix ***X*** (predictor variables) that maximises the covariance between those vectors and a matrix ***Y*** (dependent variables). PLS decomposes the matrix of zero-mean variables ***X*** and the

matrix of zero-mean variables ***Y*** into the form $\boldsymbol{X} = TP^T + E$ and $\boldsymbol{Y} = UQ^T + F$, where the T, U are matrices of p extracted scores vectors (components, latent vectors), P, Q represent matrices of loadings, and E, F are the residuals matrices.[83]

OPLS method is often applied with the aim to improve interpretation of PLS models and reduce model complexity. OPLS removes variation from ***X*** that is not correlated to ***Y*** or, in other words, removes systematic variation in ***X*** that is orthogonal to ***Y***. Therefore, one of the main advantages of OPLS is more thrifty PLS models and easier understanding. Comparing with the PLS representation of ***X***, OPLS utilizes information in the response matrix ***Y*** to further decompose the matrix ***X*** into three distinct structures $\boldsymbol{X} = T_p{P_p}^T + T_0{P_0}^T + E$, where $T_p{P_p}^T$ contains systematic covariance and correlation structures in relation to Y, $T_0{P_0}^T$ contains systematic ***Y***-orthogonal variation and the residual matrix E contains the remaining un-modeled variation.[82]

As can be seen in Figure 1.13, the PLS components cannot separate between-class variation from the within-class variation, and the resulting PLS component loadings mixes both types of variation. The component 1 ($t1_p$) of OPLS model is the predictive component and displays the between-class (black circles, white squares) variation of the samples. The corresponding loading profile can be used for identifying variables important for the class separation. Component 2 ($t2_0$) is the ***Y***-orthogonal component and models the within group variation.[82]

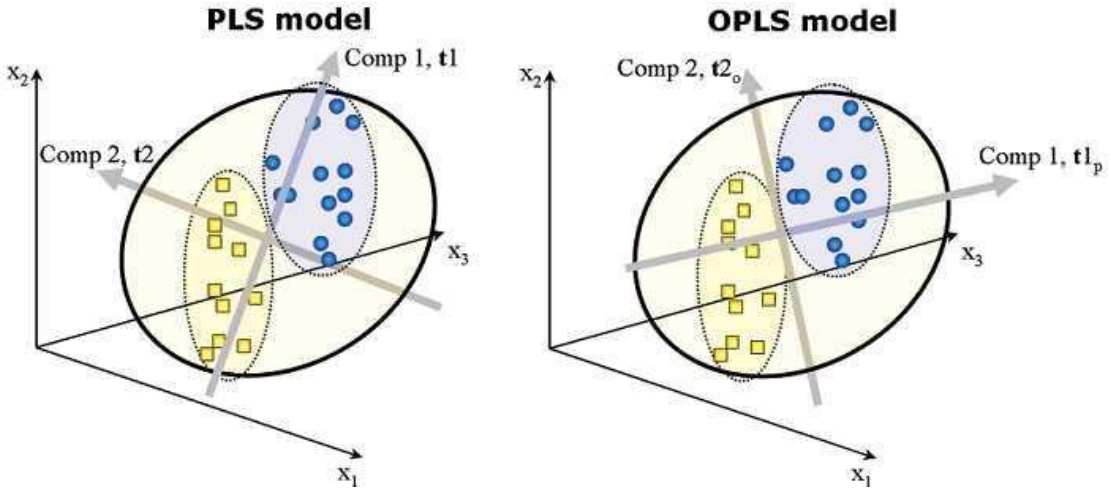


Figure 1.13 Illustration of the difference between the PLS-DA and OPLS-DA models. [Reprinted from reference [82]]

Both PLS and OPLS can be used for discrimination analysis, called PLS-DA and OPLS-DA, respectively. In this case, the ***Y*** matrix contains qualitative values as class labels. In the case of two classes, usually the values of the dependent variable are given 1 for one class and 0 for the other class.[84]

Other advanced version of PLS is interval PLS (iPLS) which is a variation of the normal PLS regression method and is employed for region selection in the information-rich ^{1}H NMR spectrum. Its main advantage is to provide an overall graphical illustration of the variation in ***X*** relevant to the dependent ***Y*** variable. In iPLS model, the NMR spectrum is divided into rather arbitrary segments, usually for equal size, and the predictive performance of each segment is assessed.[85-86]

2.1 Biological Samples

2.1.1. Collection and storage

Blood samples were obtained from pregnant women aged 17 to 41 years old, during their second trimester of pregnancy (15-24 weeks of gestation). Sample collection was carried out (from december 2005 until february 2006, and from february 2008 until march 2009), in the Prenatal Diagnosis Service in Bissaya Barreto Maternity in Coimbra within an ongoing research project (FCT/PTDC/QUI/66523/2006: An NMR-based metabonomics study of health disorders of foetus and mother during pregnancy: towards biochemical characterisation and early diagnostics).

Blood collection was carried out as follows: blood samples were collected for sodium heparin tubes, then they were centrifuged (1500xg, 10 min) and supernatants collected, frozen at -20^{o}C, and after 2 hours, stored at -70 ^{o}C until NMR analysis.

The use of the biofluids for research was subjected to ethical committee approval. Information about the occurrence of fetal and maternal health problems, intake of medication, chronic maternal disease and maternal life style was obtained from the obstetrical and neonatal medical records, as well as from an individual questionnaire, handed out at the time of sample collection. All women who participated in this study performed the amniocentesis due to risk factors such as advanced maternal age, indication of warning signs (e.g. fetal malformations) in ultrasonography, previous pregnancy with chromosomal abnormalities or positive biochemical screening.

The disorders studied in the project mentioned above were: pre-gestational diabetes (pre-GD), fetal malformations (FM), preeclampsia, intrauterine growth restriction, chromosomal anomalies, premature rupture of the membrane, and premature birth.

In this work the plasma samples studied fell into three groups: controls (healthy pregnancy), pre-GD and FM because the number of samples in the other groups was still limited at the time. The control group is composed of plasma of pregnant women who underwent a healthy pregnancy throughout the whole period. The pre-GD group was composed of plasma of pregnant women who developed the disease later in pregnancy, with clinical diagnosis usually falling between 24 and 28 weeks of gestation. The FM group is composed of plasma of pregnant women whose ultrasonography revealed several types of malformations. General information about the samples such as gestational and maternal age, as well as clinical relevant information is shown in Appendix 1.

2.1.2 Sample preparation

For NMR analysis, the samples were thawed during about 30 min at room temperature and 400 µL of saline solution (NaCl 0.9% in 10% D_2O) were added to 200 µL of plasma, D_2O being used for lock. The mixture was then centrifuged (8000 rpm, 5 min) to separate cell fragments suspended in plasma samples and transferred into 5 mm NMR tubes. No reference compound such as sodium salt of 3-trimethylsilylpropionic acid (TSP) was used due the high protein content of plasma samples and possibility of binding to proteins, thus leading to reference signal broadening.[87]

For LC-NMR and FTIR analysis, the samples were thawed for ca. 30 min at room temperature and each sample was used without pre-treatment. For LC-NMR, the volume of each blood plasma sample injected was 100 µL. For FTIR, the volume of each sample placed on the ATR crystal was 5 µL.

2.2. Data Acquisition and Processing

2.2.1 NMR spectroscopy data

The 1D and 2D NMR spectra of blood plasma samples were recorded on a Bruker Avance DRX 500 spectrometer operating at a proton frequency of 500 MHz. Sample temperature was set at 300 K and continuously controlled throughout the experiments. The sample was loaded into the probe and left at least 10 min to equilibrate to the set temperature.

For each sample, three 1D ^{1}H NMR spectra were obtained: a standard 1D, Carr-Purcell-Meiboom-Gill (CPMG), and diffusion edited spectra. Standard 1D ^{1}H NMR spectra were acquired using a noesy 1D pulse sequence (noesypr1d, RD–90º–t1–90º–tm–90º–acquire),with 100 ms mixing time (t_m), 128 scans and a receiver gain of 57. 1D CPMG spectra were acquired using the CPMG pulse sequence (cpmgpr, RD–90º–{τ–180º–τ}n–acquire) with simple presaturation of the water peak and a total spin-spin relation time ($2n\tau$) of 240 ms, 256 scans and a receiver gain of 128. 1D diffusion edited spectra were recorded using the bipolar pulse longitudinal eddy current delay (BPPLED) pulse sequence, with 128 scans, and a receiver gain of 32. In order to attenuate the signals from low molecular weight compounds, gradient duration of 2 ms and strength of 48.24 G/cm were used, with 100 ms diffusion time. All 1D spectra were acquired with 32 k complex data points, 8012.82 Hz spectral width (SW), 5 s relaxation delay and 4 s dummy scans. For data processing, each FID was zero-filled to 64 k points, multiplied by a 0.3 Hz exponential line-broadening function prior to Fourier transform. Spectra were baseline corrected for phase and baseline distortions using TOPSPIN 2.1 and referenced internally to the anomeric proton signal of α-glucose at 5.23 ppm.

TOCSY spectra were acquired using time proportional phase increment (TPPI) mode, and the MLEV17 pulse sequence was used for spin lock. 64 FIDs were acquired with 2048 complex data points per increment to a total of 128 increments. Spectral width of 8012.820 Hz were employed in both dimensions, together with a 111ms mixing time, and a 2s relaxation delay. HSQC spectra was acquired with inverse detection, ^{13}C decoupling during acquisition. 32 FIDs were recorded with 4096 complex data points per increment up to a total of 300 increments. Spectral widths of 8012.820 and 25153.811 were used for ^{1}H and ^{13}C dimensions, respectively. JRES spectra were acquired using a total of 32 FIDs with 16384 complex data points, and spectral width of 8012.820 in F2; in F1 (J-coupling) 64 data points were acquired, with a spectral width of 15.654.

Diffusion-ordered spectroscopy (DOSY) experiments were acquired using the bipolar pulse longitudinal eddy current delay (BPPLED) pulse sequence. Typically, in each Pulsed Field Gradient-NMR experiment, a series of 16-64 BPPLED spectra of 16-32 k data points were collected. The values of δ and Δ were 2.1 ms and 250 ms duration, respectively, and

the eddy current delay (Te) was set to 5 ms in all experiments. The pulse gradients (g) were incremented from 12 to 95% of the maximum gradient strength in a linear ramp. After Fourier transformation and baseline correction the diffusion dimension was processed with Bruker TOPSPIN software package (Version 2.1).

The assignments of 1D and 2D 1H NMR spectra were achieved by comparison of spin systems with those found in a Bruker Spectral Database and in literature data (see Appendix 1 for detailed assignment of plasma ^{1}H NMR spectrum).[70]

2.2.2 Liquid Chromatography-Nuclear magnetic Resonance Spectroscopy (LC-NMR)

Potassium phosphate buffers for LC were prepared using potassium di-hydrogen phosphate (KH_2PO_4) and potassium phosphate di-basic 3-hydrate ($K_2HPO_4 \cdot 3H_2O$) salts purchased from Panreac, and deuterium oxide purchased from Euriso-top. The LC column was packed using Tilesius-type hydroxyapatite Bio-Gel HTP (DNA grade) purchased from Bio-Rad Laboratories.

The HPLC equipment consisted of an Agilent 1100 solvent delivery system with a quaternary pump and solvent degasser (Agilent, Waldbronn, Germany), a manual injector equipped with a 100 µL sample loop from Rheodyne, and a diode-array detector (DAD) from Bruker Biospin (Rheinstetten, Germany). The Agilent 1100 system was connected through a BPSU-36/2 interface (Bruker Biospin, Rheinstetten, Germany) to an NMR Avance spectrometer, working at 500.13 MHz for proton, with LC-SEI probe head (120 µL active volume). Blood plasma (100 µL) was injected onto a Tilesius-type hydroxyapatite Bio-Gel HTP column (12 x 0.46 cm). A flow rate of 0.15 mL/min was used, and a gradient of potassium phosphate buffer + 10% $^{2}H_2O$ (pH 7.4) employed with stepwise changes from an initial concentration of 75 mM to 200 mM after 30 min and to 650 mM after 60 min. DAD detection was set at 280 nm and 200-300 nm.

Along the chromatographic runs, fractions were stored by loop collection and eventually transferred to the probe head of NMR in order to acquire the ^{1}H NMR spectra.

The ^{1}H NMR spectra were recorded at 500.13 MHz using a probe temperature of 300K. The “lc1pnf2” pulse sequence from Bruker library with double solvent presaturation was employed with DQD acquisition mode, 128 transients, 32k data points, a spectral width of 8012.820 Hz and 2s acquisition time. Each FID was zero-filled to 64k, multiplied by an exponential function corresponding to a line broadening of 0.5 Hz, manually phased and baseline corrected. Chemical shifts for the HPLC NMR-derived lipoprotein spectra were referenced internally to the resonance from the residual acetonitrile at δ 2.07 ppm existing inside the probe head.

2.2.3 NMR spectral deconvolution

The information about the number of metabolites and chemical shift obtained by LC-NMR was used to perform deconvolution on three different regions of standard 1D and diffusion-edit ^{1}H NMR spectra: terminal methyl protons (δ 0.76-1.06), long chain methylene protons (δ 1.14-1.36), phospholipid choline headgroups (δ 3.16-3.28). These regions were fitted to Lorentzian lineshapes using MestReNova (Version 6.0.4, Mestrelab Research). The resonances included in each spectral region are given in Table 2.1. After deconvolution, the peak areas of HDL, LDL, VLDL, and saturated fatty acids were used to calculate the relative percentage of each component using $\% = (A_i \times 100) \div \sum_{n=i} A$, where A is the area of each compent denoted by *i* (HDL, LDL, VLDL, saturated fatty acids).

Table 2.1 Number of peaks and resonances considered in deconvolution of three spectral regions on standard 1D ^{1}H NMR spectrum of a control sample (PS 11406) at 15 weeks of gestational age.

Region	Component	δ_H (ppm)	Height	Width (Hz)	Area
δ 0.76 – 1.06 ppm	HDL	0.83	2.86E+07	19.48	3.30E+09
	LDL	0.85	2.66E+07	16.21	3.52E+09
	VLDL	0.86	3.00E+07	18.07	3.71E+09
	Saturated fatty acids	0.89	1.00E+07	24.27	3.28E+09
	Isoleucine, Leucine	0.93	1.20E+07	36.91	2.25E+09
	Isoleucine, Leucine	0.95	1.28E+07	36.31	2.25E+09
	Valine	0.98	7.89E+06	5.02	2.25E+08
	Valine	0.99	8.27E+06	5.02	2.18E+09
	Isoleucine	1.01	1.09E+07	14.74	1.14E+09
	Valine	1.03	1.04E+07	5.37	1.64E+09
	Valine	1.04	1.30E+07	5.72	1.57E+09
δ 1.14 – 1.36 ppm	HDL	1.23	2.94E+07	24.60	2.16E+09
	LDL	1.25	2.39E+07	22.75	2.33E+09
	VLDL	1.27	3.04E+07	13.79	2.62E+09
	Saturated fatty acids	1.29	1.57E+07	17.59	1.92E+09
	Threonine	1.31	8.65E+06	16.22	1.34E+09
	Lactate	1.31	1.58E+07	2.67	3.75E+09
	Threonine	1.32	6.49E+06	11.16	1.18E+09
δ 3.16 – 3.28 ppm	Choline	3.194	2.87E+06	9.17	2.88E+08
	Choline	3.209	9.08E+06	8.77	7.68E+08
	Choline	3.218	1.28E+07	8.27	8.71E+08
	Glucose	3.221	6.93E+06	2.27	3.50E+09
	Choline	3.229	1.86E+06	5.71	4.10E+08
	Glucose	3.237	4.84E+06	2.27	5.51E+08
	Glucose	3.241	6.34E+06	2.27	7.54E+08
	Arginine	3.248	2.32E+06	5.71	2.50E+08
	Glucose	3.256	7.58E+06	2.27	4.17E+08

2.2.4 FTIR spectroscopy data

FT-IR measurements were performed on a FT-IR Perkin Elmer spectrometer, Spectrum BX model equipped with an Attenuated Total Reflectance (ATR) Golden Gate Specac. The spectra were obtained in the wavenumber range of 600-4000 cm^{-1} with a spectral resolution of 4 cm^{-1}, 1 cm^{-1} of interval and 64 scans.

Two procedures for drying the sample were tested for reproducibility and rapidity: drying at room temperature (24±0.4ºC and 33±3%) and drying with a stream of nitrogen. In both methods, a background spectrum was acquired (a measure with no sample in the crystal which allows to remove all instrumental characteristics), 5 μL of intact plasma were positioned onto the centre of the diamond anvil and one spectrum was immediately acquired. Then, for drying at room temperature, spectra were acquired with 10 min between until the sample was completely dry. For drying with a stream of nitrogen spectra were acquired with 5 minutes between them. The sample was considered to be dry when there were no visible differences in intensity between three consecutive spectra.

Upon comparison of both methods, the drying at room temperature was chosen for analysis of all samples because it showed better reproducibility between consecutively acquired spectra. To test the reproducibility between samples acquired in different days, spectra of three aliquots of five control samples were used and PCA was used to better visualize spectral reproducibility.

The procedure used to acquired all samples was: spectra acquired with drying at room temperature with 10 minutes between them until the sample is completely dry (about 30 min), and, after this, five consecutive spectra were acquired. Thus, for each sample five spectral replicas were obtained after drying. The aim of using five FTIR spectra for each sample was to increase the number of points in ***X*** matrix for MVA. The room temperature and humidity was 24±0.6 ºC and 35±3 %.

The assignment of the FTIR spectra of plasma was performed using information from literature (Table 1.3).[74]

2.3. Chemometrics Analysis

2.3.1. NMR data analysis

Each set of spectra (noesy 1D, CPMG and diffusion-edited) was used to construct data matrices for the MVA, either using the full spectra or spectral regions. The aim of using discrete spectral regions is to decrease the number of peaks to be analysed in each model in order to perform a more detailed analysis. Thus, six spectral regions covering all relevant signals and excluding the water spectral region (δ 4.5-5.0) were considered for separate analysis (Table 2.2). In the case of full spectra, all signals in δ 0.5-10.0 region except the water spectral region were included in analysis.

Table 2.2 Spectral regions used in MVA of NMR data

Spectral Regions (ppm)	General Assignment
δ 0.5 – 1.4	CH_3, $(CH_2)_n$ of lipids and amino acids
δ 1.4 – 3.1	Amino acids, lipids, organic acids
δ 3.1 – 3.3	$-N^+(CH_3)_3$ choline and glucose
δ 3.3 – 4.5	Glucose and amino acids
δ 5.0 – 5.5	Mannose, glucose and lipids
δ 5.5 – 10.0	Amino acids

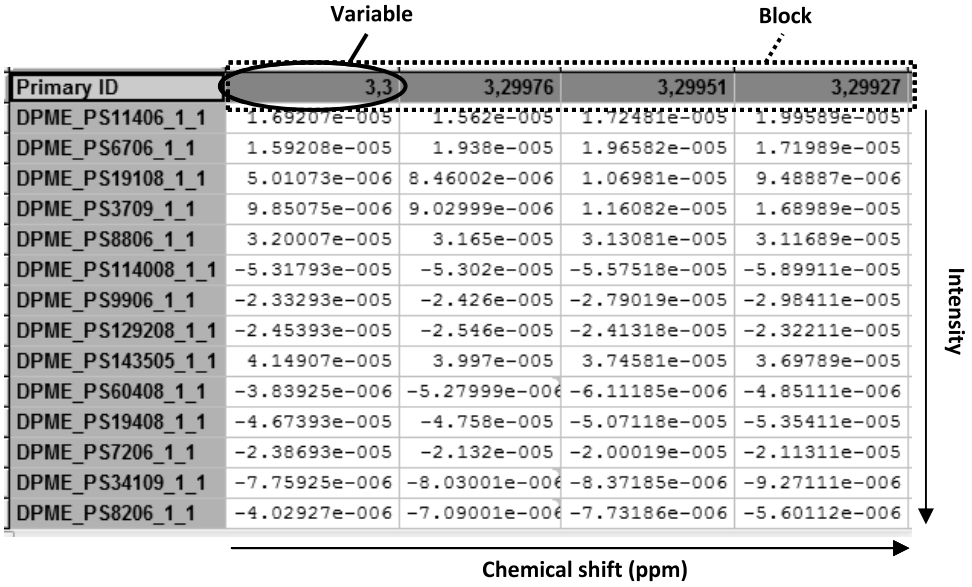
Each spectrum was exported from AMIX-Viewer version 3.9.1 in two distinct ways: divided into buckets (or bins) of 0.01 ppm and full resolution. In both cases, spectra were normalized to total intensity. The aim of bucketing spectra is to reduce the effect of pH-induced changes in chemical shift, ensuring that the same species are always counted correctly across samples with such variation. In addition, normalization is the adjustment of the spectral intensities with the aim to make the data from all samples comparable with each other (each value of a vector is divided by the square root of the sum of the squares of the elements of this vector).

The resulting data were imported into SIMCA-P (version 11.5.0.0; Umetrics) for MVA. PCA, PLS-DA and OPLS-DA were performed on no scaling data and two scaling conditions: UV and pareto. Scaling was performed on each spectral intensity across all samples and its aim is to give unit variance.

In Figure 2.1 can be seen an illustration of a NMR data matrix where samples are on the rows and spectral intensities for each ppm value are on the columns. Considering each ppm value as a variable and all ppm values as a block, both unscaled and scaled

data were mean centered in which each value of a given variable is subtracted by the mean value of this variable with the aim to place all values in the same reference axis. UV scaling divides each mean centered variable by its standard deviation, i. e., gives equal variance to all variables of a block. Pareto scaling is in between no scaling and UV scaling and divides each mean centered variable by the square root of its standard deviation, i. e., gives to all variables a variance equal to their standard deviation instead of unit variance.

In fact, when MVA is applied to scaled data, the resulting model takes into account each peak intensity in the same manner, independently of the fact it representing major or minor peaks on the ^{1}H NMR spectra.



Primary ID	3,3	3,29976	3,29951	3,29927
DPME_PS11406_1_1	1.69207e-005	1.562e-005	1.72481e-005	1.99589e-005
DPME_PS6706_1_1	1.59208e-005	1.938e-005	1.96582e-005	1.71989e-005
DPME_PS19108_1_1	5.01073e-006	8.46002e-006	1.06981e-005	9.48887e-006
DPME_PS3709_1_1	9.85075e-006	9.02999e-006	1.16082e-005	1.68989e-005
DPME_PS8806_1_1	3.20007e-005	3.165e-005	3.13081e-005	3.11689e-005
DPME_PS114008_1_1	-5.31793e-005	-5.302e-005	-5.57518e-005	-5.89911e-005
DPME_PS9906_1_1	-2.33293e-005	-2.426e-005	-2.79019e-005	-2.98411e-005
DPME_PS129208_1_1	-2.45393e-005	-2.546e-005	-2.41318e-005	-2.32211e-005
DPME_PS143505_1_1	4.14907e-005	3.997e-005	3.74581e-005	3.69789e-005
DPME_PS60408_1_1	-3.83925e-006	-5.27999e-006	-6.11185e-006	-4.85111e-006
DPME_PS19408_1_1	-4.67393e-005	-4.758e-005	-5.07118e-005	-5.35411e-005
DPME_PS7206_1_1	-2.38693e-005	-2.132e-005	-2.00019e-005	-2.11311e-005
DPME_PS34109_1_1	-7.75925e-006	-8.03001e-006	-8.37185e-006	-9.27111e-006
DPME_PS8206_1_1	-4.02927e-006	-7.09001e-006	-7.73186e-006	-5.60112e-006

Figure 2.1 Illustration of a NMR data matrix.

Besides the methods mentioned above, an extended version of PLS-DA was applied, called interval PLS-DA (iPLS-DA). It was employed for region selection in the spectra and works by sweeping the NMR data to search variations relevant to the classes. Data divided into buckets of 0.01 ppm was used. iPLS-DA was applied to spectra divided into 20 segments, and the predictive performance (robustness) of each segment was assessed.

After the MVA analysis, the criteria used to evaluate the results were:

1. Inspection of scores plot to see how the points were distributed in PC space. All results were visualized by two-dimensional scores plots which represent the distribution of samples in the model.

2. If the scores plot showed any trend between the two classes, the validity of the model was evaluated. The parameters used to evaluate the validity were the R^2X, R^2Y and

Q^2. For PLS-DA and OPLS-DA models, R^2X is the explained variance of X explained, and R^2Y is the explained variance of Y. Ideally, the ratio between R^2X and R^2Y should be small because a small R^2X value and a high R^2Y value indicates that, in the construction of the model, the class was more relevant than the ^{1}H NMR spectra. Finally, Q^2 value is the cross validated R^2 and can be used to test the validity of the model; higher Q^2 values are usually associated with best discrimination between the two classes.

3. The loadings plots were observed to known which metabolites were responsible for the trend. The loading plots provide information on the contribution of each variable to the pattern in the scores plots. For improved interpretation, loadings plots of spectral regions with many peaks were plotted using for each point a color corresponding to the weight value in the model that represents the correlation of the *X* variable (NMR spectra) with *Y* class. The interpretation of the loadings is therefore straightforward because the resulting plot provides a loading with the same shape as that of a spectrum but, in addition, on the same plot, the variable importance in the projection (VIP) for the discrimination between the classes are highlighted by the color code.

4. In order to validate the results obtained by MVA, metabolites that contribute to the trend between the two classes were integrated using AMIX-Viewer version 3.9.1 and normalized to total spectral area, excluding water resonance, can be compared using boxplots. A boxplot is a simple univariate tool that detects outliers and that can compare distributions of the data among different groups. The separation between the two classes can be after evaluated by statistical tests. First, shapiro-wilk normaly test was applied to determine if the data followed a normal distribution or not. For normally distributed data, the t-student test was then applied, and for not normally distributed data, Wilcoxon test was used. The p-value obtained by the statistical test allows the confirmation or not of the differences between the two groups. In this study, the null hypothesis is “average differences are equal to zero” and the alternative hypothesis is “average differences are not equal to zero”. For p-values lower than 0.05, the alternative hypothesis is confirmed which means that the integrals are significantly different between the two classes.

5. The percentage of variance of each important metabolite in each prenatal disease sample relative to controls was determined by $\% = ((A_{metabolite} - \bar{x}_C)/\bar{x}_C) \times 100$, where, $A_{metabolite}$ is the normalized area of each metabolite in each prenatal disease sample, and $\bar{x}_C$ is the mean area of the control group.

Figure 2.2 provides an overview of the strategic adopted in the study of ^{1}H-NMR-based metabolic profile of maternal blood plasma for pre-GD and FM groups.

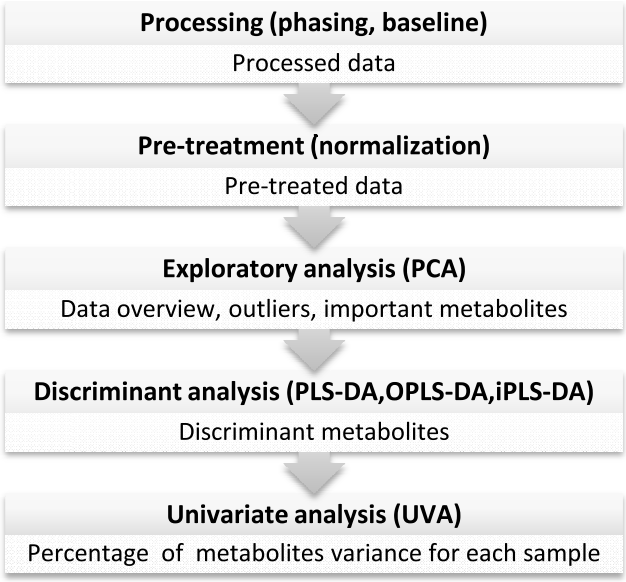


Figure 2.2 Strategy adopted in the study of ^{1}H-NMR-based metabolic profile of maternal blood plasma.

2.3.2. FTIR data analysis

The plasma FTIR data used for MVA consisted of five spectral replicas per sample acquired after drying for 30 min. The resulting data was transferred to Cats (chemometrics analysis tool system), a program developed by the University of Aveiro and Institut National Agronomique Paris-Grignon.[88]

The FTIR data was used to construct data matrices, either using the full spectra (4000-600 cm^{-1}) without CO_2 absorbance region (2400-2200 cm^{-1}) or spectral regions. In the later case, four spectral regions covering relevant absorbance bands were considered for separate analysis as can be seen in Table 2.3.

Table 2.3 FTIR spectral regions chosen for chemometrics analysis

Spectral Regions (cm^{-1})	Assignment
3110-2800	Fatty acids
1800-900	Mixture of biomolecules
1800-1475	Amide I and amide II of proteins
1200-900	Glucose, lactate, glycerol

Full absorbance spectra and spectral regions were converted to first and second derivative absorbance spectra. The aim of considering spectra derivative functions is to minimize the intensity differences in spectra in order to maximize the variations between

the classes. Thus, variations between the groups that could be undetectable in normal absorbance spectra can be detected using this manipulation technique.

Thus, three distinct data matrices were constructed: absorbance spectra, first derivative absorbance spectra and second derivative absorbance spectra. The derivatives were performed in Cats and using the Savitzky-Golay algorithm.

Data matrices were pre-processed by centering and standardizing rows (spectra) and centering columns (wavenumber). Data standardization involves that each value of a given variable is divided by the standard deviation value of this variable. Then, these matrices were imported into SIMCA-P (version 11.5.0.0; Umetrics), where PCA and PLS-DA were performed. As in the results obtained of NMR MVA, the results were visualized by two-dimensional scores plots and the corresponding loadings plots, and the values of R^2X, R^2Y and Q^2 were considered to assess the predictive performance of each model as well as for NMR data.

3.1. Study of Prenatal Disorders by ^{1}H NMR Spectroscopy and LC-NMR

3.1.1. Assignment and Interpretation of NMR spectra of Human Blood Plasma

As mentioned before, the assignment of the 1D ^{1}H NMR spectra of complex samples like blood plasma is very complex. Hence, several 1D and 2D ^{1}H NMR spectra were acquired in order to assign the major metabolites of human blood plasma. The assignment published by Nicholson et al.[70] was the first very useful account of assignments in this work.

The standard 1D, CPMG and diffusion-edited ^{1}H NMR spectra of blood plasma of a healthy pregnant woman at 18 weeks of gestation are represented in Figure 3.1. The standard 1D ^{1}H NMR spectrum (Figure 3.1a) is dominated by the broad resonances of large molecules such as lipoproteins and proteins present in blood plasma as well as by many signals from low M_w metabolites. Signals from the N-acetyl groups of glycoproteins and lysyl groups of albumin are detected. A high number of signals arise from the various types of protons in the fatty acyl chains of lipoproteins also appear such as terminal CH_3 and long chain CH_2 protons of fatty acids, and head groups of the choline-containing phospholipids in lipoproteins. Furthermore, several resonances from small metabolites can also be detected in the standard 1D spectrum, such as glucose, some amino acids (valine, alanine, arginine, threonine), lactate, acetate, acetoacetate, pyruvate, citrate, creatine, creatinine, and formate.

The CPMG spectrum, shown in Figure 3.1b, allows the study of low M_w metabolites above, since the broad signals from large molecules are attenuated by the CPMG pulse sequence. The signals from several low M_w compounds already identified by standard 1D spectrum can be identified such as amino acids, organic acids and glucose. Also, resonances from isoleucine, histidine, aspartate, mannose and methionine, only appear in CPMG spectrum. Some vestigial peaks from lipoproteins are already detected in CPMG spectrum such as the terminal CH_3 and long chain CH_2 protons of fatty acids, and choline containing phospholipids.

The diffusion-edited ^{1}H NMR spectrum (Figure 3.1c) uses the difference in molecular diffusion coefficients between macromolecules and low M_w metabolites to attenuate narrow peaks, which is especially useful for the study of lipids and lipoprotein particles. Compared to the standard 1D spectrum, the peaks arising from fast diffusing metabolites are strongly attenuated, while the broad resonances corresponding to slowly diffusing macromolecules are not affected. Hence, the spectrum in figure is dominated by the characteristic lipidic resonances, mainly the terminal CH_3 and long chain CH_2 protons of fatty acids in lipoprotein particles.

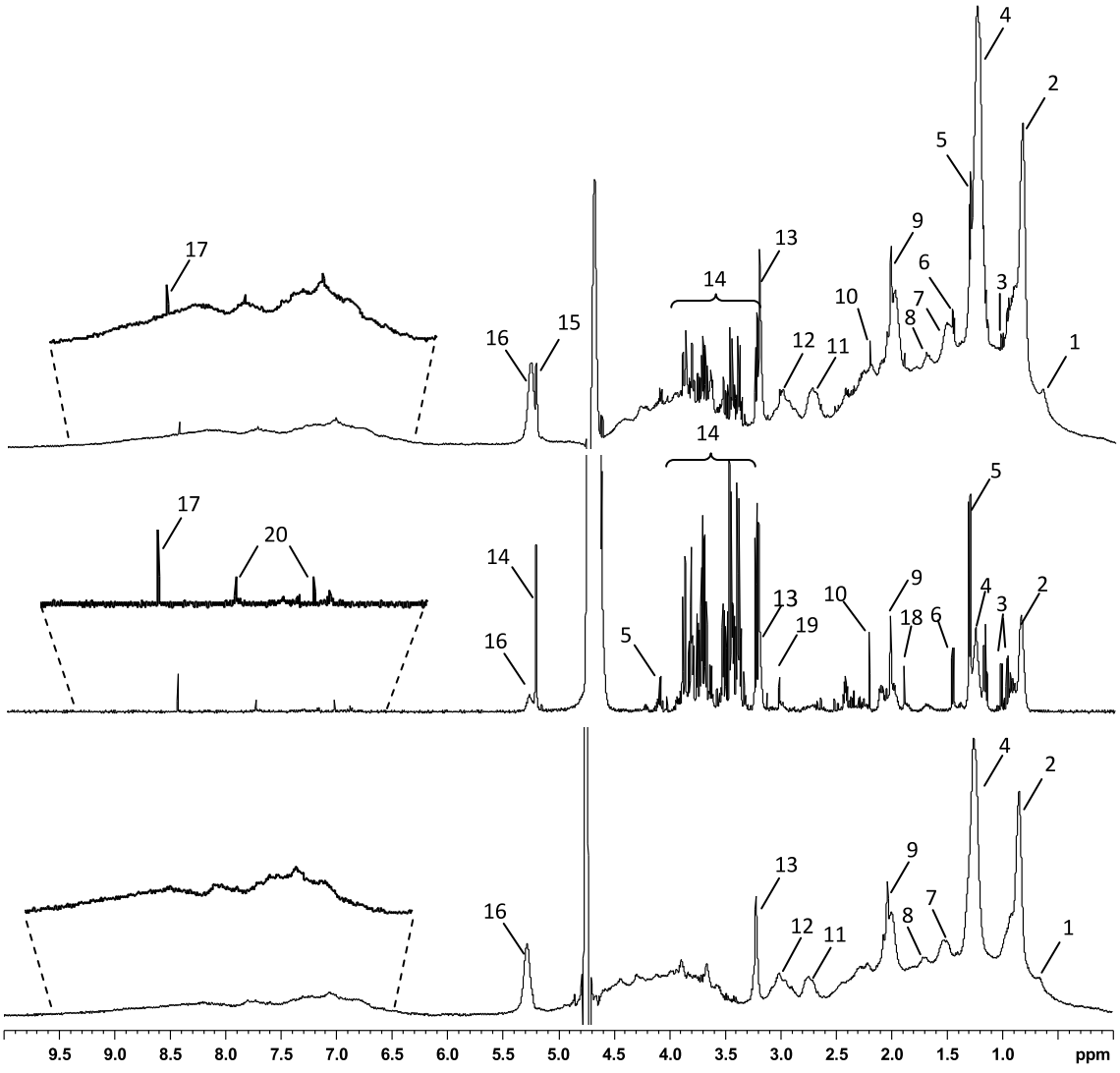


Figure 3.1 500 MHz ^{1}H NMR spectra (δ 0-10) of blood plasma of a healthy pregnant woman (PS 10306) at 18 weeks of gestational age, measured at 300 K. (a) standard 1D spectrum; (b) CPMG spectrum; and (c) diffusion edited spectrum. Assignment: 1, cholesterol C18; 2, lipid C$\underline{H}_3$; 3, valine; 4, lipid (C$\underline{H}_2$)$_n$; 5, lactate; 6, alanine; 7, lipids C$\underline{H}_2$CH$_2$CO; 8, lipid C$\underline{H}_2$CH$_2$C=C; 9, glycoprotein NHCOCH$_3$; 10, acetoacetate; 11, lipids =CHC$\underline{H}_2$CH=; 12, albumin lysyl ε-CH$_2$; 13, phospholipids choline headgroup –N$^+$(CH$_3$)$_3$; 14, α- and β-glucose; 15, α-glucose; 16, lipids -CH=CH-; 17, formate; 18, acetate; 19, creatine and creatinine; 20, histidine.

In order to improve the assignment of the ^{1}H NMR spectra of plasma, several 2D experiments were acquired. The ^{1}H-^{1}H TOCSY spectrum of blood plasma for the same healthy pregnant woman is shown in Figure 3.2. This spectrum is dominated by major macromolecular lipoprotein components such as terminal CH$_3$ and long chain CH$_2$ protons of fatty acids, and protons from C$\underline{H}_2$C=C, =CC$\underline{H}_2$C=, phospholipid choline headgroups N$^+$(CH$_3$)$_3$, and HC=CH. Low molecular weight metabolites can also be assigned such as

some amino acids, lactate and glucose. Lysine, glutamine and glyceryl groups of lipids can only be assigned in this experiment.

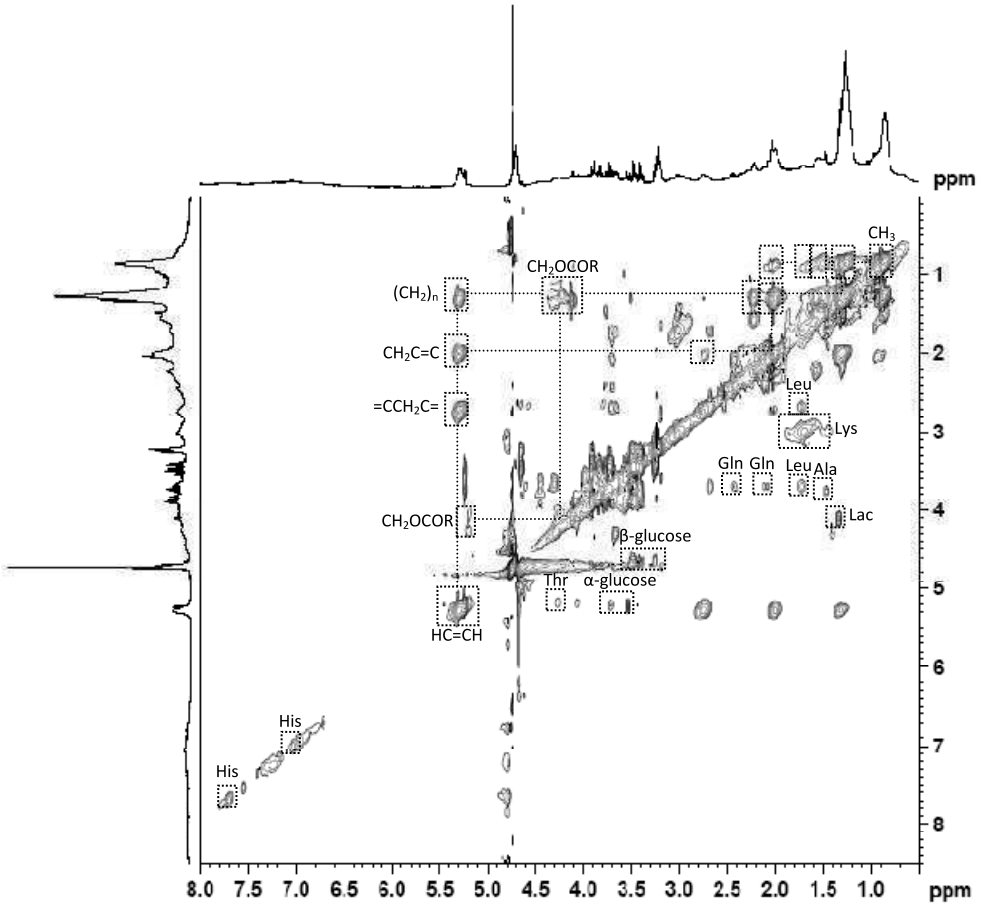


Figure 3.2 500 MHz ^{1}H-^{1}H TOCSY of blood plasma from the same healthy pregnant woman (PS 10306) at 18 weeks of gestational age, measured at 300 K. Abbreviations: Ala, alanine; His, histidine; Lac, lactate; Leu, leucine; Lys, lysine; Gln, glutamine; Thr, threonine.

The signal overlap persists in ^{1}H-^{1}H TOCSY spectrum and peaks from other components of plasma are not observed because of their weak intensities relative to lipid signals. Compared with other ^{1}H-^{1}H TOCSY spectra of plasma found in literature, the information is very similar with lipidic resonances from lipoprotein to dominate the spectrum.[89]

Figure 3.3 shows the ^{1}H-^{13}C HSQC of blood plasma for the same sample. This 2D spectrum allowed the more detailed assignment of lipid resonances such as CH_3 groups of cholesterol, the long chain CH_2 protons of fatty acids and $N^{+}(CH_3)_3$ groups of choline-containing phospholipids among others. The signals from acetyls of glycoproteins and all of the α- and β-glucose resonances can be assigned, as well as the resonances of lysyls of

albumin (ϵCH_2) which can't be assigned in 1D ^{1}H NMR spectra. The correlation spectra found in literature[70] also provides strong peaks for lipidic resonances and the ^{13}C chemical shifts are very similar to that obtained in this work. An important disadvantage of this type of experiments is the relatively long acquisition times, as mentioned before due the low isotopic abundance of ^{13}C.

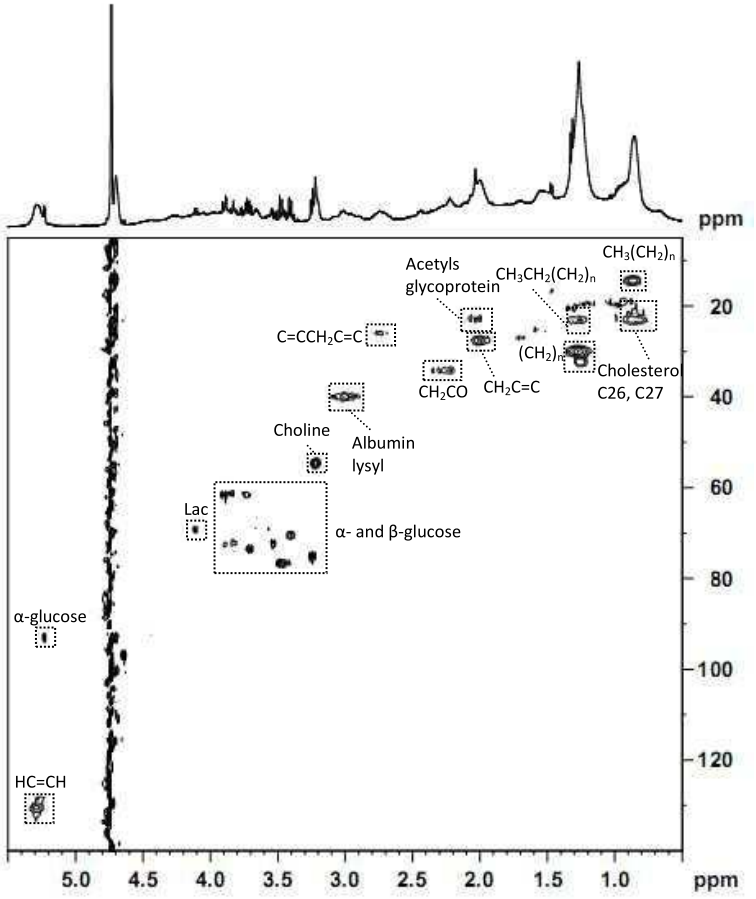


Figure 3.3 500 MHz ^{1}H-^{13}C HSQC of blood plasma from the same healthy pregnant woman (PS 10306) at 18 weeks of gestational age, measured at 300 K. Abbreviations: Lac, lactate.

Finally, JRES experiment was also acquired for the same sample to provide complementary information to the TOCSY and HSQC. Figure 3.4 shows signals from lipoproteins attenuated and the complex overlapped resonances in the chemical shift range from δ 3-4 are better resolved (Figure 3.4b). The metabolite peaks assigned in JRES spectrum are labelled in the figure. Several signals previously assigned in 1D ^{1}H NMR

spectra were identified and the resonance from αCH group of glycine was only observed in this experiment. The total list of assignments obtained is presented in Table 3.1.

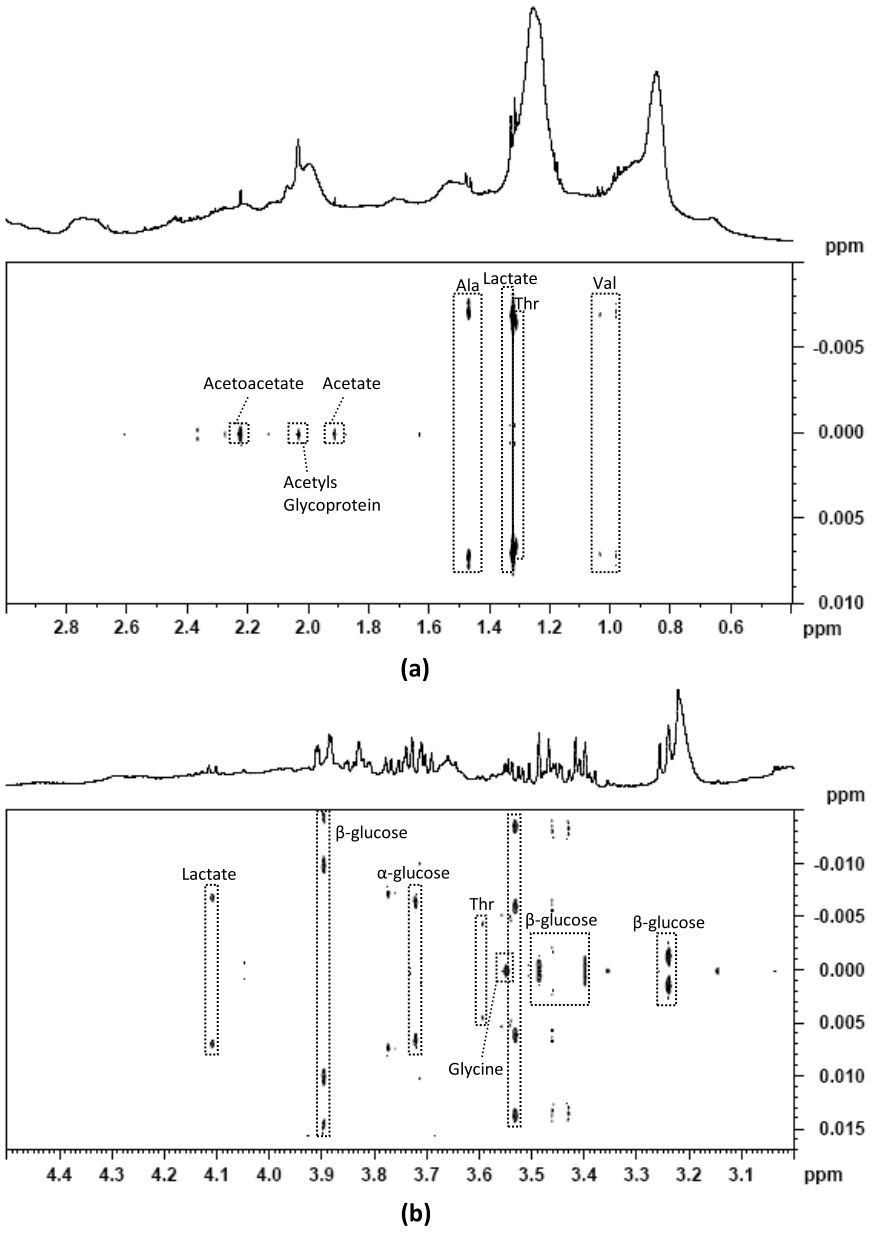


Figure 3.4 500 MHz ^{1}H JRES NMR spectrum of blood plasma from the same healthy pregnant woman (PS 10306) at 18 weeks of gestational age measured at 300 K. (a) Spectral region covering δ 3.0-0.4 and (b) spectral region covering the range δ 3.0-4.5.

Table 3.1 500 MHz ^{1}H NMR assignments for human blood plasma of a healthy pregnant woman (PS 10306) at 18 weeks of gestational age, and pH around 8.0. Abreviations: s, singlet; d, doublet; dd, doublet of doublets; t, triplet; q, quartet; m, multiplet.

Compound	δ_H, ppm (multiplicity, assignment) / δ_C, ppm	NMR experiment
Acetate	1.91 (s,CH_3)	Standard 1D, CPMG, JRES
Acetoacetate	2.22 (s,CH_3)	Standard 1D, CPMG, JRES
N-acetyl groups (glycoproteins)	2.03 (s,CH_3)/22.7	Standard 1D, CPMG, diffusion-edited, HSQC
Alanine	1.46 (d,CH_3)	Standard 1D, CPMG, JRES, TOCSY
Albumin lysyl	2.90 (t, ϵCH_2)/39.86; 2.96 (t, ϵCH_2)/39.86; 3.01 (t, ϵCH_2)/39.95	Standard 1D, diffusion-edited, HSQC
Arginine	1.68 (m, γ CH_2)	Standard 1D, CPMG
Aspartate	2.68 (dd, β CH_2)	CPMG
Citrate	2.51 (d, α,β,CH); 2.66 (d, α',β',CH)	Standard 1D, CPMG
Cholesterol	0.65 (m, C18H); 0.84 (m, C26,27H)/22.95	Standard 1D, diffusion-edited, HSQC
Choline	3.21 (s, $N(CH_3)_3$)/54.6	Standard 1D, CPMG, diffusion-edited, TOCSY, HSQC
Creatine	3.03 (s, NCH_3); 3.92 (s, NCH_2)	Standard 1D, CPMG
Creatinine	3.04 (s, NCH_3); 4.05 (s, NCH_2)	Standard 1D, CPMG
Formate	8.45 (s, CH)	Standard 1D, CPMG
α-glucose	3.41 (t, C4H)/70.34; 3.53 (dd, C2H)/72.19; 3.71 (t, C3H)/73.49; 3.76 (dd, 5,8, C6H)/61.51; 3.83 (m, C5H)/72.01; 3.83 (m, C6H')/61.22; 5.23 (d, 3,9, C1H)/92.87	Standard 1D, CPMG, TOCSY, HSQC
β-glucose	3.23 (dd, C2H)/74.97; 3.40 (t, C4H)/70.34; 3.47 (dd, C5H); 3.49 (t, C3H)/76.65; 3.71 (dd,C6H')/61.55; 3.89 (m, C6H)/61.51; 4.64 (d, 8,1, C1H)/96.71	Standard 1D, CPMG, TOCSY, HSQC
Glutamine	2.08 (m, βCH_2); 2.10 (m, $\beta' CH_2$); 2.42 (γ CH_2); 3.69 (t, αCH)	TOCSY
Glyceryl of lipids	4.1 (m, CH_2OCOR); 4.22 (m, CH_2OCOR); 5.19 (m, CHOCOR)	TOCSY
Glycine	3.55 (s, αCH)	JRES
Histidine	7.04 (s, C4H, ring); 7.74 (s, C2H,ring)	CPMG, TOCSY
Isoleucine	0.95 (t, γCH_3); 1.00 (d, βCH_3)	CPMG
Lactate	1.32 (d, CH_3); 4.10 (q, CH)/69.23	Standard 1D, CPMG, TOCSY, HSQC
Leucine	0.95 (d, δCH); 0.96 (d, δCH); 1.69 (m, γCH); 1.71 (m, βCH_2); 3.70 (dd, αCH)	CPMG, TOCSY
Lipid	0.85 (m, CH_3); 1.25 (m, $CH_3CH_2CH_2$)/32.17; 1.26 (m, $CH_3CH_2(CH_2)_n$)/22.98, 1.27 (m, $(CH_2)_n$)/29.91, 32.1; 1.53 (m, $\underline{CH_2}CH_2CO$); 1.70 (m, $\underline{CH_2}CH_2C{=}C$); 1.99 (m, $\underline{CH_2}C{=}C$)/27.52; 2.74 (m, C=C $\underline{CH_2}C{=}C$)/25.8; 5.29 (m, HC=CH)/130.2	Standard 1D, CPMG, diffusion-edited, TOCSY, HSQC
Lysine	1.69 (m, δCH_2); 2.99 (t, ϵCH_2);	TOCSY
Mannose	5.18 (d, αCH)	CPMG
Methionine	2.13 (m, βCH_2); 2.14 (s, SCH_3)	CPMG
Pyruvate	2.36 (s, CH_2)	Standard 1D, CPMG
Threonine	1.32 (dd, β'CH); 3.60 (d, αCH); 4.24 (m, βCH)	Standard 1D, CPMG, JRES, TOCSY
Valine	0.98 (d, γCH_3); 1.03 (d, $\gamma' CH_3$); 2.26 (m, βCH)	Standard 1D, CPMG
Unassigned spin system	1.41 (d)	Standard 1D, CPMG

As can be seen by the several 1D and 2D NMR spectra showed above, the assignment of the lipoprotein resonances to the individual lipoprotein classes is not possible. This is because, as expected, the different lipoprotein classes (HDL, LDL and VLDL) show very similar and overlapping ^{1}H NMR signals due to their similar chemical composition. However, the assignment of the lipoprotein particles can be very useful for clinical diagnosis mainly to provide information on the composition without physical or chemical separation. Thus, in order to attempt a characterization of lipoprotein particles, the diffusion ordered spectroscopy (DOSY) was applied to blood plasma from a healthy pregnant woman at 16 weeks of gestation, as can be seen in Figure 3.5.

However, interpretation of the DOSY spectrum is very difficult because of the extensive signal overlap. The diffusion-weighted separation between low molecular weight metabolites, proteins and lipids is unclear. Nevertheless, there is a tendency for low molecular weight metabolites to the top part of the spectrum (high diffusion coefficients), and proteins and lipids for the bottom part of spectrum (low diffusion coefficients). The assignment of low molecular metabolites is limited to the glucose, lactate and some amino acids resonances. The aromatic proton region from δ 6.4 to 8.6 ppm contains several resonances from amino acids, however, the assignment is not possible.

The assignment of the individual lipoprotein signals is also difficult due the very small differences in diffusion coefficients of lipoprotein particles. An expansion of DOSY spectrum in the region of methyl and methylene resonances of lipoproteins is also shown in Figure 3.5. The diffusion coefficient within the chemical shift region δ 0.84-1.29 ppm ranging between 2.40×10^{-11} (-10.62 log (m^2/s)) and 4.57×10^{-11} m^2/s (-10.34 log (m^2/s)) at 300K.In the literature, the diffusion coefficients of lipoproteins and some low M_w metabolites in blood plasma can be found. This knowledge was possible by use of 2D diffusion-edited total-correlation NMR spectroscopy (DETOCSY) at 37°C. However, the assignment to each lipoprotein class was also not possible. For lipoprotein resonances, the average of diffusion coefficients was ~1×10^{-10} m^2/s.[90] These values of diffusion coefficients can not be directly compared with those obtained due to the differences in experimental temperature. At 37°C, the plasma viscosity is lower and, hence, the motional freedom of the lipoprotein particles is higher than at 27°C. Thus, the diffusion coefficients for lipoprotein particles measured at 37°C are higher than at 27°C.

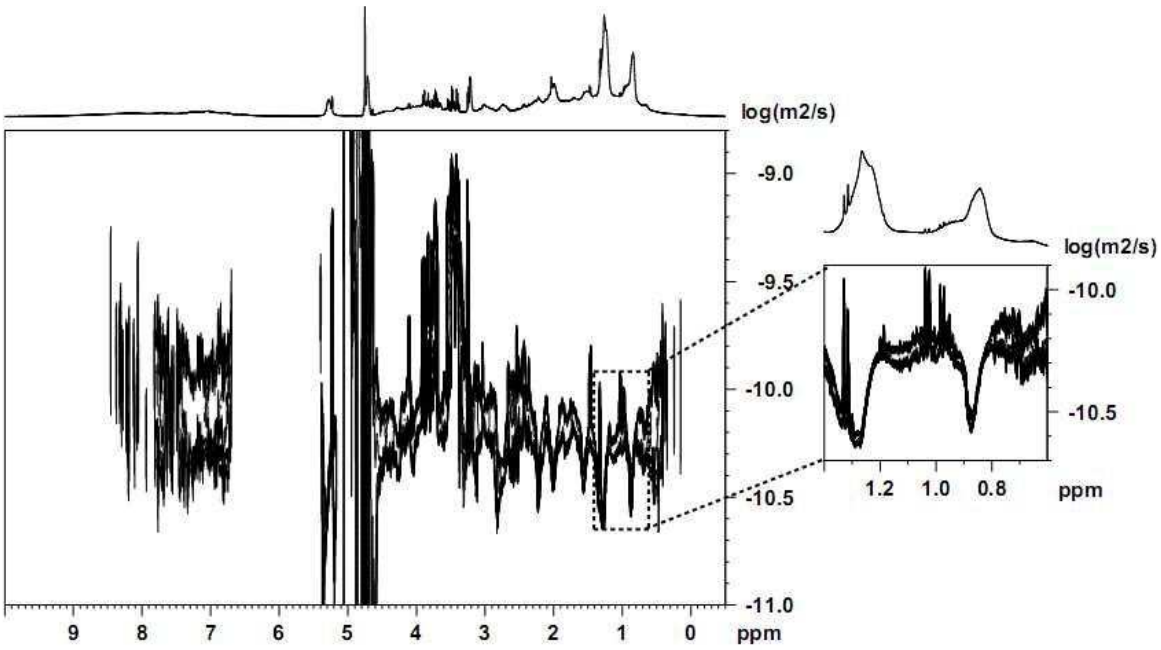


Figure 3.5 500 MHz ^{1}H DOSY NMR spectrum of blood plasma from a healthy pregnant woman (PS 35808) at 16 weeks of gestation measured at 300K, with spectral expansion covering δ 0.75-1.4.

3.1.2. LC-NMR for Separation and Characterization of Blood Lipoproteins

In this work, the method used by Daykin *et al.* to separate the three major lipoprotein classes (HDL, LDL, VLDL) was reproduced.[36] The separation of each lipoprotein class provides information about their chemical shift profile and peak width, which can be used to perform mathematically deconvolution of the heavily overlapped methyl, methylene and choline resonances in ^{1}H NMR spectra of blood plasma. The intention was to further characterize the lipoprotein signals, in order to be able to quantify these components directly in whole blood plasma, without previous separation.

Due to experimental problems caused by some technical defaults in the LC-NMR apparatus, only three samples were studied: two control and one pre-GD samples. Slight variations in lipoprotein composition were observed among the three samples analysed, however, some variability is expected since the samples could not be collected under fasting conditions and are, therefore, affected by diet. This can be a limitation and the only way to overcome this problem and obtain consistent comparison between controls and diseased subjects is to analyse a large number of samples, in the future. At this point, however, the information obtained on the limited number of samples was employed to aid direct analysis of whole plasma samples.

The UV detection chromatogram (280 nm) for a control sample (PS 11406) is shown in Figure 3.6 with ten fractions collected during the chromatographic procedure.

According to the original work published by Daykin *et al.*, the three peaks on the UV chromatogram correspond to each of the major lipoprotein class, HDL, LDL and VLDL, respectively.[36]

The composition of each fraction collected in this work is given by the ^{1}H NMR spectra in Figure 3.7a., and they can be compared with those obtained by Daykin *et al.* for HDL, LDL and VLDL, in Figure 3.7b.

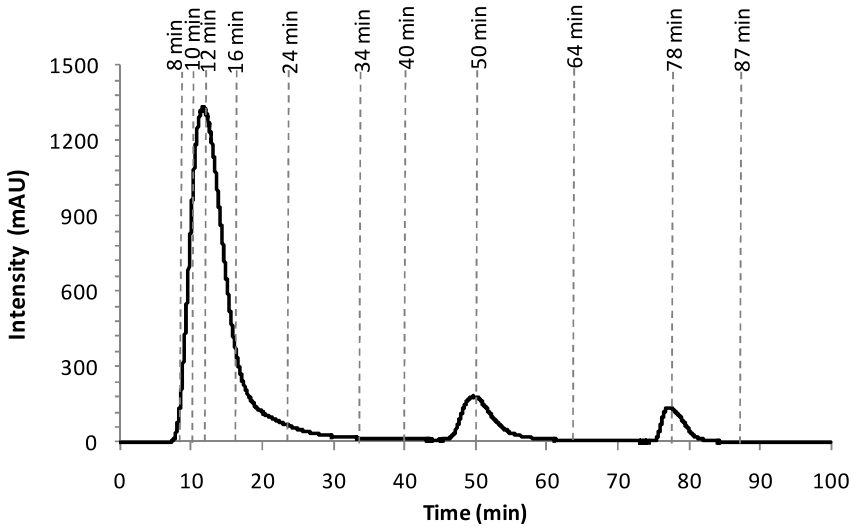


Figure 3.6 UV-detected chromatogram at 280 nm of a healthy pregnant woman (PS 11406) blood plasma.

As can be seen in Figure 3.7a, at 8 min, only residual compounds were eluted. From 10 to 16 min, the ^{1}H NMR spectra is very similar to the standard 1D spectra of whole plasma, with broad resonances from proteins together with narrow resonances from low M_w metabolites such as glucose, lactate, and amino acids. The chemical shifts of the terminal CH_3 and the CH_2 groups of fatty acids appear at δ 0.83 and δ 1.23 ppm, respectively. Other characteristic lipidic resonances can be seen, such as the C18 methyl groups of cholesterol and adjacent methylene protons in $C\underline{H}_2CH_2CO$ side chains of lipids. The phospholipids choline headgroup $N^+(CH_3)_3$ shows three types of environments (or compounds), as can be seen by the three shoulders in figure. These fractions have spectra similar to that of the HDL fraction shown in Figure 3.7b in terms of retention time and chemical sifts of the terminal CH_3 and the CH_2 groups of fatty acids, however, shows more low M_w metabolites.

From 24 min to 40 min, only residual resonances, probably from proteins and lipoproteins, can be seen. The ^{1}H NMR spectrum at 50 min shows only characteristic lipidic resonances, the terminal CH_3 and the CH_2 groups of fatty acids appear at δ 0.84 and δ 1.25 ppm, respectively. In this fraction only one resonance from phosphoplipids choline

headgroup $N^+(CH_3)_3$ can be viewed at δ 3.213 ppm. This fraction is very similar in terms of chemical shifts and retention time to LDL fraction seen in Figure 3.7b.

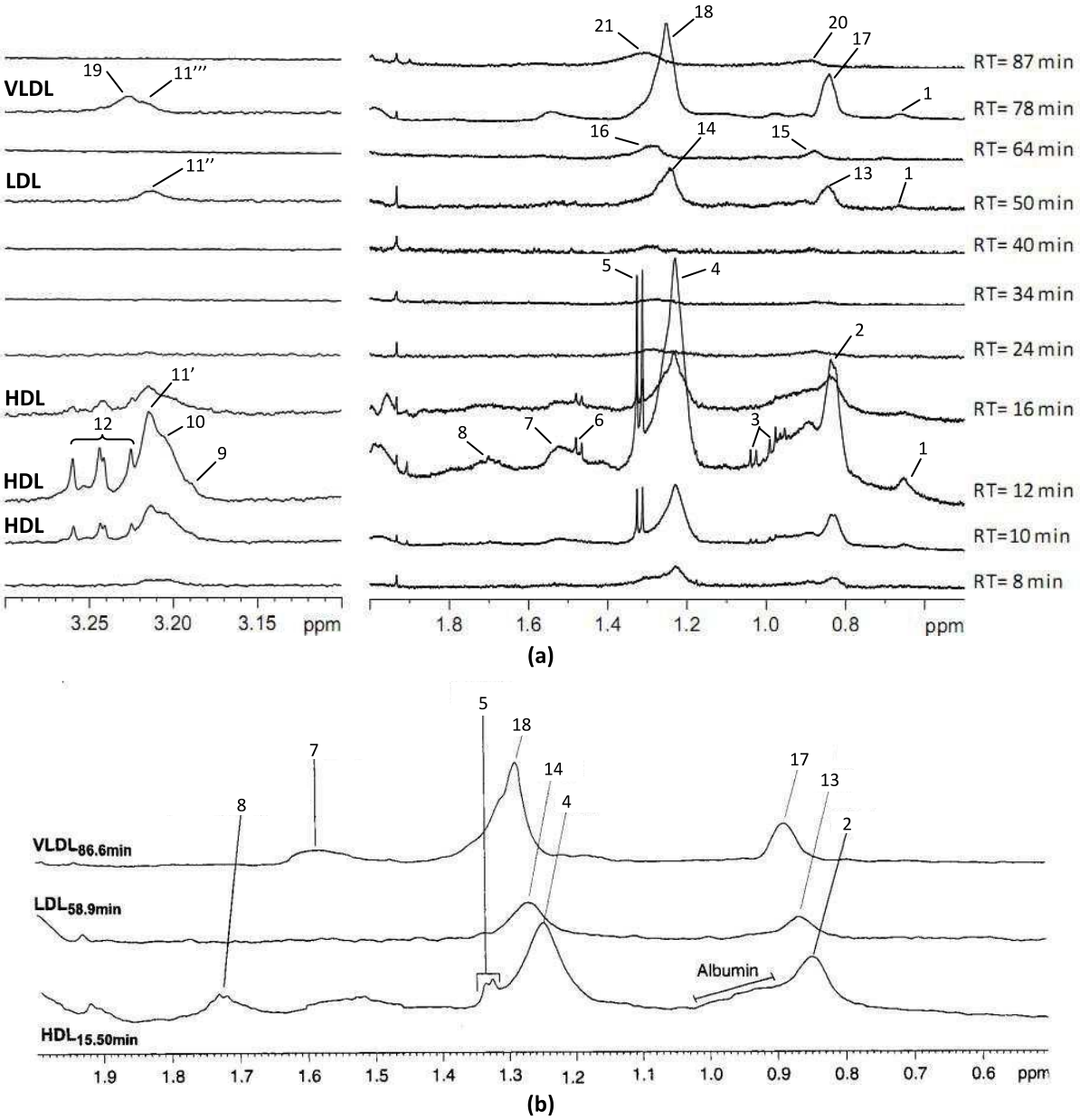


Figure 3.7 Comparison and assignment of the fractions collected by LC-NMR in this work and literature[36]. (a) δ 3.1-3.3 and δ 0.5-2.0 expansion of 500 MHz LC ^{1}H NMR spectra. Assignment: 1, cholesterol C18; 2, HDL CH_3; 3, valine; 4, HDL $(CH_2)_n$; 5, lactate; 6, alanine; 7, lipids CH_2CH_2CO; 8, lipids $CH_2CH_2C{=}C$; 9, 10 and 11', HDL $N^+(CH_3)_3$; 11'', LDL $N^+(CH_3)_3$; 11''', VLDL $N^+(CH_3)_3$; 12, glucose; 13, LDL CH_3; 14, LDL $(CH_2)_n$; 15, saturated fatty acids CH_3; 16, saturated fatty acids $(CH_2)_n$; 17, VLDL CH_3; 18, VLDL $(CH_2)_n$; 19, VLDL $N^+(CH_3)_3$; 20, saturated fatty acids CH_3; 21, saturated fatty acids $(CH_2)_n$. (b) 600 MHz LC ^{1}H NMR spectra (δ 0.5-2.0) of the HDL, LDL and VLDL separated by LC-NMR at 15.5, 58.9 and 86.6 min, respectively.

At 64 min, saturated fatty acids appear shifted, with generally higher chemical shifts compared to the fractions collected at 50 and 78 min. These lipid compounds may either occur in the free form in plasma or be derived from the lipoprotein breakdown.

The ^{1}H NMR spectrum obtained for 78 min shows several lipidic resonances, the terminal CH_3 and the CH_2 groups of fatty acids appearing at δ 0.84 and δ 1.25, respectively. Two resonances from choline headgroup $N^{+}(CH_3)_3$ can be visualized at δ 3.213 and δ 3.225 ppm. This fraction can be assigned to VLDL shown in Figure 3.7b. In the last fraction collected at 87 min, saturated fatty acids can be observed.

In summary, the characteristic terminal CH_3 and the CH_2 groups of fatty acids in fractions collected from 10 to 16 min, at 50, and at 78 min show chemical shift profiles consistent with those reported in literature[36] for HDL, LDL and VLDL, respectively. In Table 3.2, the comparison of chemical shifts and retention times obtained in this work and the literature can be seen.

Table 3.2 Comparison of the chemical shifts and retention obtained by LC-NMR of blood plasma in this work and the informationfound in the literature. SM, sphingomyelin; PC, phosphatidylcholine.

Class	This work			Literature		
	RT/min	δ/ppm (Peak*)	Assignment	RT/min	δ/ppm (Reference)	Assignment
HDL	10-16	0.83 (2)	CH_3	15-22	0.83 (36)	CH_3
		1.23 (4)	$(CH_2)_n$		1.23 (36)	$(CH_2)_n$
		3.19 (9)	$N^{+}(CH_3)_3$		3.21 (91)	SM $N^{+}(CH_3)_3$
		3.21 (10)	$N^{+}(CH_3)_3$			
		3.21 (11')	$N^{+}(CH_3)_3$		3.22 (91)	PC $N^{+}(CH_3)_3$
LDL	50	0.84 (13)	CH_3	58	0.84 (36)	CH_3
		1.25 (14)	$(CH_2)_n$		1.25 (36)	$(CH_2)_n$
		3.21 (11'')	$N^{+}(CH_3)_3$		3.24 (91-92)	SM $N^{+}(CH_3)_3$
					3.25 (91-92)	PC $N^{+}(CH_3)_3$
Saturated Fatty Acids	64	0.88 (15)	CH_3			
		1.28 (16)	$(CH_2)_n$			
VLDL	78	0.85 (17)	CH_3	87	0.85 (36)	CH_3
		1.26 (18)	$(CH_2)_n$		1.26 (36)	$(CH_2)_n$
		3.21 (11''')	$N^{+}(CH_3)_3$			
		3.23 (19)	$N^{+}(CH_3)_3$			
Saturated Fatty Acids	87	0.89 (20)	CH_3			
		1.30 (21)	$(CH_2)_n$			

* peak number in Figure 3.7a

The phospholipid choline headgroup resonances numbered as 11', 11'' and 11''' have the same chemical shifts, however, they are present at very different retention times, thus arising from distinctly different compounds. As can be seen in Table 3.2, the resonances of choline-containing phospholipids in blood plasma have assigned to sphingomyelin and phosphatidylcholine in HDL and LDL classes.[91-92] The chemical shifts of sphingomyelin and phosphatidylcholine are similar to choline resonances numbered as 10 and 11', respectively, in HDL class. However, for LDL class only one choline compound can be viewed, and the chemical shift is different from the phospholipids found in literature.

The deconvolution of the standard 1D ^{1}H NMR spectra of blood plasma was then accomplished using the information obtained by LC-NMR and the shoulders visualized in the spectral bands. For PS 11406 sample, the deconvolution of the methylene and terminal methyl of fatty acids, and choline-containing phospholipids regions (δ 0.76-1.06, δ 1.14-1.36, δ 3.16-3.28 ppm, respectively) are presented in Figure 3.8 where the blue curves are the single components considered and the lineshape in red is the difference between the experimental and fitted spectra. Other non-lipidic resonances (e. g. leucine, isoleucine, valine) had to be considered in order to obtain an accurate deconvolution of the lipids, as described in detail in the experimental section (2.2.3). The accuracy of the deconvolution can be evaluated by the red line in figure.

In order to calculate the percentage of HDL, LDL, VLDL and saturated fatty acids, the spectral region used was the terminal methyl resonance of fatty acids at δ 0.76-1.6 ppm, since it is the most commonly used signal area in lipoprotein analysis and the total number of terminal methyl groups contained in each class has lesser sources of variability in the relative amounts than the long chain methylene groups.[39, 93] Thus, using the terminal methyl peak areas for HDL, LDL, VLDL and saturated fatty acids, the percentage of each spin system was calculated, as well as, the percentages of the different choline containing compounds resonating at δ 3.16-3.28 ppm, the results are shown in Table 3.3.

The percentages of HDL, LDL, VLDL and saturated fatty acids obtained for a control sample (PS 11406) at 15 weeks of gestational age were 24, 25, 27 and 24% respectively. For choline-containing compounds, the peaks at δ 3.194 and δ 3.209 ppm only seen in HDL, had percentages of 12 and 33%, respectively. The choline compound resonating at δ 3.218 ppm observed in the three lipoprotein classes had a percentage of 37%. Finally, for choline compound resonating at δ 3.229 ppm only observed in VLDL, the percentage obtained was 18%.

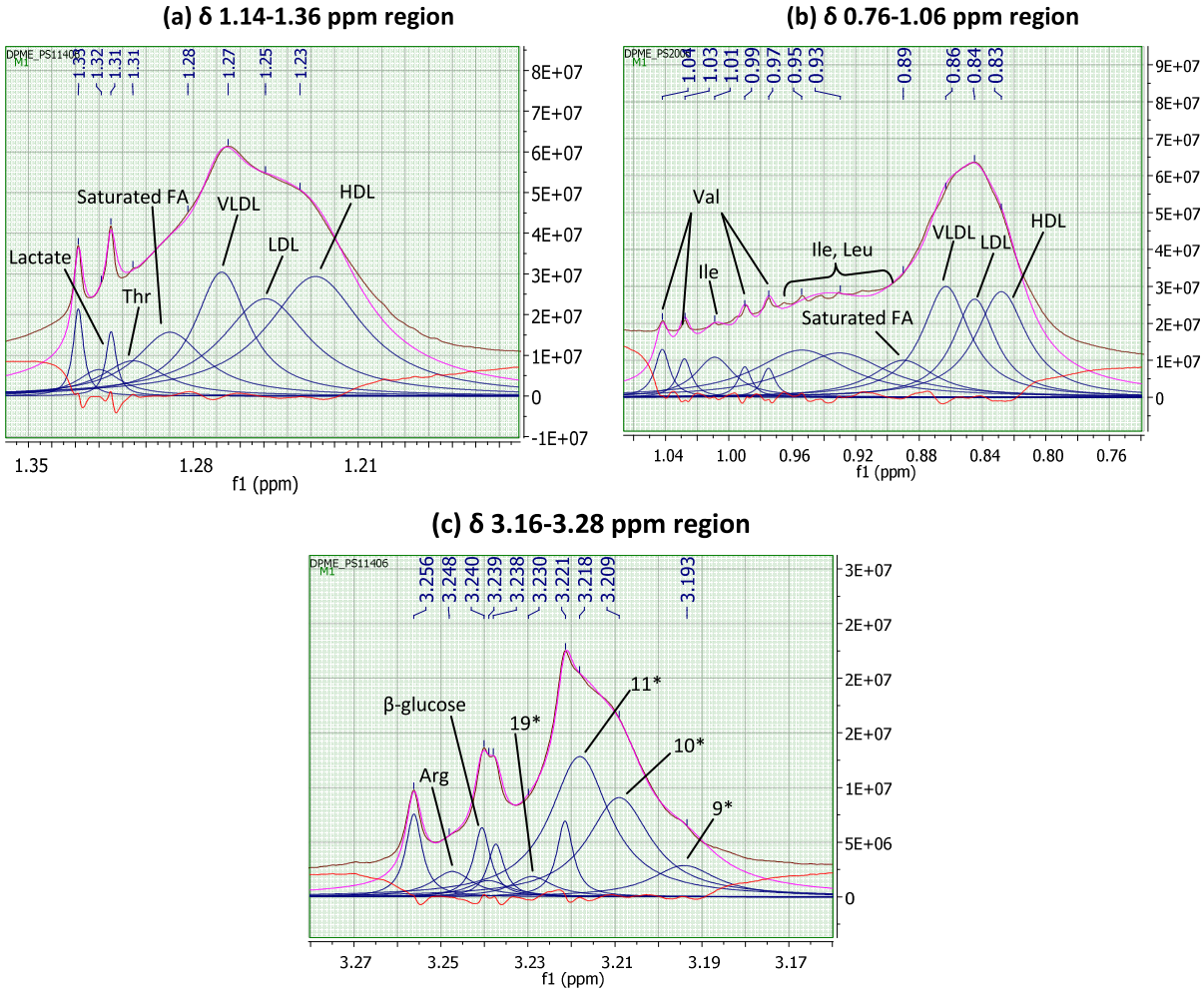


Figure 3.8 Spectral deconvolution applied to the standard 1D ^{1}H spectrum of blood plasma from a healthy pregnant woman (PS 11406) at 15 weeks of gestation. (a) and (b) long chain fatty acid methylene and terminal methyl resonances at δ 1.14-1.36 and δ 0.76-1.06 ppm, respectively; (c) phospholipids choline headgroup $N^{+}(CH_3)_3$ resonances at in δ 3.16-3.28 ppm. The blue lines are the single components considered, the pink line is the simulated profile of the band, and the red line is the difference between experimental and simulated profile of the band. Abbreviations: Arg, arginine; FA, fatty acids; Ile, isoleucine; Leu, leucine; Thr, threonine; Val, valine.

Table 3.3 Percentages of lipoprotein classes and choline-containing compounds calculated from results obtained by deconvolution of standard 1D ^{1}H NMR spectrum of a control sample (PS 11406) at 15 weeks of gestation. FA, fatty acids.

Description	Resonance	δ_H (ppm)	%
Lipoprotein classes (δ 0.76 – 1.06 ppm)	HDL CH_3	0.83	24
	LDL CH_3	0.85	25
	VLDL CH_3	0.86	27
	Saturated FA CH_3	0.89	24
Choline-containing compounds (δ 3.16 – 3.28 ppm)	9* $N^{+}(CH_3)_3$	3.19	12
	10* $N^{+}(CH_3)_3$	3.21	33
	11* $N^{+}(CH_3)_3$	3.21	37
	19* $N^{+}(CH_3)_3$	3.23	18

* peak number in Figure 3.7a

3.1.3. Gestational Diabetes Viewed by NMR

In order to search for predictors of GD, MVA was applied to the ^{1}H NMR profile of plasma of women with pre-GD (n=12) and healthy pregnant women (n=20). As mentioned before, pre-GD group is composed by women who developed GD later in pregnancy, so that the time of samples collection preceded the time of GD diagnosis, as can be seen in Table 3.4. It is recalled that the aim of this study was to search for biomarkers of GD present in maternal metabolism before the diagnosis, which may allow the prediction of the disease and, hence, improve prenatal care.

Table 3.4 Pre-GD samples considered in the study and respective gestational age at sample collection and at pre-GD diagnosis. W, weeks; d, days.

Samples	Gestational age at sample collection	Gestational age at GD diagnosis
PS 6306	17 w	-
PS 8306	17 w	34 w + 1 d
PS 19108	17 w + 1 d	25 w + 1 d
PS 19408	17 w + 3 d	27 w + 2 d
PS 59108	17 w + 1 d	-
PS 60408	16 w + 5 d	22 w + 5 d
PS 65808	17 w + 1 d	28 w + 1 d
PS 114008	16 w + 3 d	-
PS 129208	16 w + 5 d	31 w + 3 d
PS 3709	17 w + 2 d	26 w + 3 d
PS 34109	21 w	32 w + 2 d
PS sc109	16 w	32 w

As described in the experimental section, PCA, PLS-DA and OPLS-DA models were developed by each of the three different 1D ^{1}H NMR experiment. Due the high number of models performed, the results shown are only for models with significant grouping trends between the pre-GD and control groups in the scores plots.

Firstly, it is important to show the effects of scaling in the spectra and corresponding models. An example of the PCA scores plot obtained for non-scaled and UV-scaled data of pre-GD and control samples, using the standard 1D ^{1}H NMR spectra, is shown in Figure 3.9. In both PCA scores plots, the pre-GD samples are overlapped with control samples. However, the PCA of unscaled data (Figure 3.9a) shows slightly higher dispersion than with UV-scaled data (Figure 3.9c). Comparing the explained variance, the principal component 1 (PC1) explains more variance (74%) for unscaled data than for scaled data (44%). In both plots, the same outlier (the PS 114008 sample) is observed. The PS 114008 shows a highest intensity resonance from methylene protons of fatty acids in lipoprotein particles in the ^{1}H NMR spectrum. In clinical information, this patient was in risk of

thromboembolism which can explain the high peak intensity in methylene protons of lipoproteins, since venous thromboembolism has been associated with higher levels of small dense LDL cholesterol.[94]

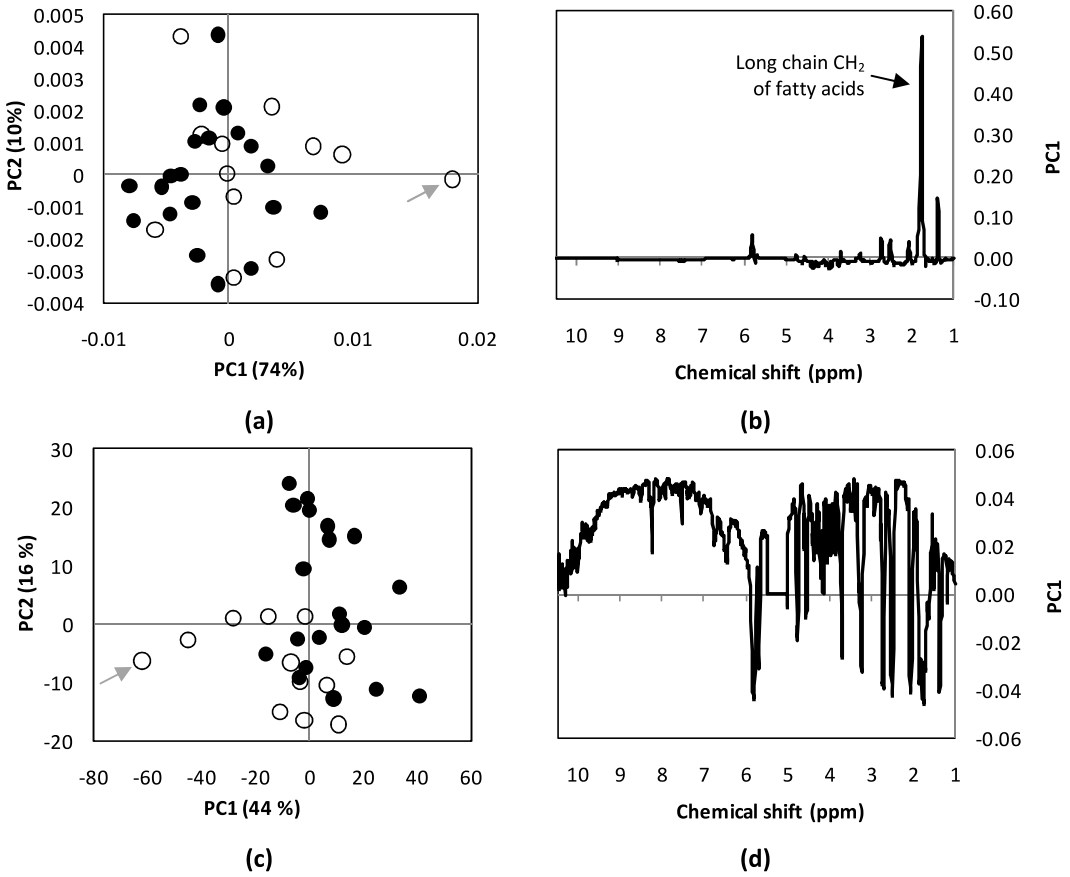


Figure 3.9 PCA models for unscaled data and UV-scaled data of control (black symbols, n=20) and pre-GD (open symbols, n=12) samples, in δ 0.5-10.0 region of the full ^{1}H NMR spectra of blood plasma. (a) and (b) PCA scores plot and loading plot of unscaled data, respectively; (c) and (d) PCA scores plot and loading plot of UV-scaled data, respectively. Outlier (PS 114008) is indicated by arrows.

Observing the PC1 loading, it is clear that unscaled data (Figure 3.9b) gives more weight to peaks with highest intensity in the ^{1}H NMR spectra, as for example the long chain methylene protons of fatty acids. On the other hand, UV-scaling (Figure 3.9d) gives equal weight to each peak, allowing small systematic changes to be more readily detected. However, UV-scaling can sometimes pick up to much noise and care must be take regarding this.

PLS-DA was then applied to unscaled data as a discrimination procedure to improve sample classification. A model with two latent variables (LVs) was built and the corresponding PLS-DA scores (Figure 3.10a) shows a slight grouping trend, however, three

strong outliers (PS 114008, PS 19408 and PS sc109) are noted (arrows in figure). The PS 114008 was already an outlier in the PCA model. In addition, the PS 19408 sample also shows higher levels of fatty acids and the PS sc109 sample has been picked up due to higher intensity peak of the ethanol peaks (δ 1.18, δ 3.65 ppm). Interestingly the clinical information of both these samples did not show abnormal characteristics.

The OPLS-DA model was also applied, the main expected benefit being the ability of OPLS-DA to separate predictive from non-predictive (orthogonal) variation (i. e., from random effects like diet). However, the OPLS-DA scores plot obtained (Figure 3.10b) did not improve on the results (both have Q^2=0.3).

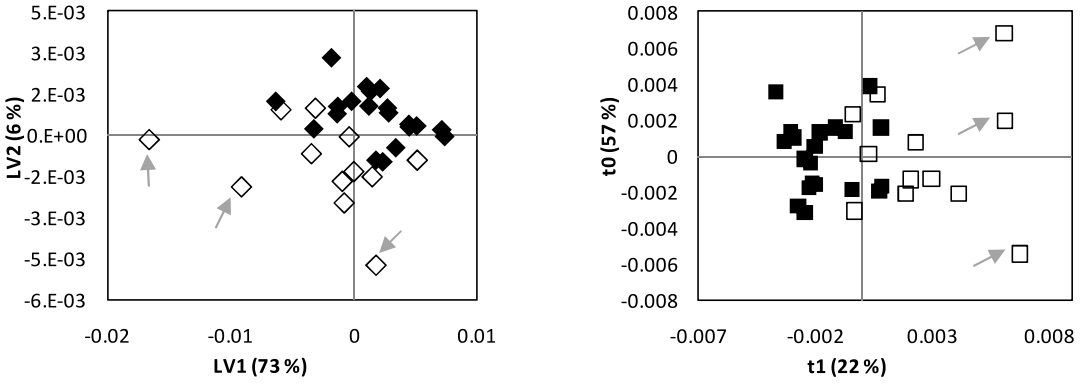


Figure 3.10 MVA of control (black symbols, n=20) and pre-GD (open symbols, n=12) samples in δ 0.5-10.0 ppm region, as viewed by standard 1D ^{1}H NMR spectra of plasma. (a) PLS-DA scores plot, LV=2, R^2X=0.79, R^2Y=0.56, Q^2=0.30; (b) OPLS-DA scores plot, LV=1+1, R^2X=0.79, R^2Y=56, Q^2=0.29. Outliers are indicated by arrows.

As described in the experimental section, each of the three different ^{1}H NMR spectra obtained for each sample (standard 1D, CPMG and diffusion-edited ^{1}H NMR spectra) was divided into six spectral regions with the aim of reducing the larger number of spectral points compared to samples. The δ 3.1-3.3 ppm region was the only one showing significant grouping trends. In the standard 1D ^{1}H NMR spectra of blood plasma this region contains β-glucose resonances (double of doublets) overlapped with choline-containing phosphoplipids resonances (singlet), as can be seen in Figure 3.11. The average spectra of controls (Figure 3.11b) and the pre-GD group (Figure 3.11c) exhibit some differences in the phosphoplipids choline resonance (rows in the figures). This indicates that a difference exists regarding choline-compounds composition.

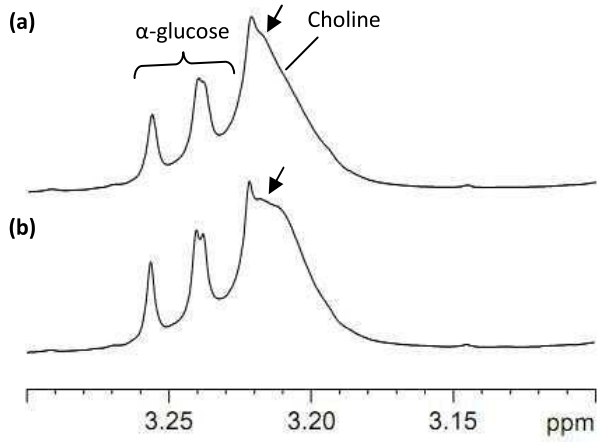


Figure 3.11 Standard 1D ^{1}H NMR average spectrum of plasma in δ 3.1-3.3 region of (a) controls and (b) pre-GD group.

The corresponding PCA (Figure 3.12a) shows a clear separation between control and pre-GD samples along PC2, except for PS 34109 sample, which is closer to the controls. However, besides pre-GD diagnosis, no other abnormal information was found in the clinical history of this sample. The PLS-DA scores plot (Figure 3.12b) was built with two LVs and a clear separation between control and pre-GD samples is seen along LV1, again except for PS 34109. This was confirmed by OPLS-DA model (Figure 3.12c) and in terms of models validity, the values of R^2X, R^2Y and Q^2 were good for both PLS-DA and OPLS-DA, characterizing them as valid models.

The OPLS-DA loadings plot (Figure 3.12d) shows the phospholipid choline at δ 3.209 ppm, previous observed by LC-NMR, and glucose in positive LV1, while the phospholipid choline resonance at δ 3.219 and δ 3.223 ppm, also previous observed in LC-NMR, are in negative LV1. These results are in accordance to PCA and PLS-DA loading plots (not shown) and indicate that pre-GD group have higher levels of phospholipid choline resonance at δ 3.209 ppm and glucose when compared to control group, and lesser contents of phospholipid choline resonance at δ 3.219 and δ 3.223 ppm. For the sake of clarity, the different choline-containing compounds will be named as: a) choline A (δ 3.209 ppm, present in HDL fraction), (b) choline B (δ 3.219 ppm, present in HDL,LDL and VLDL classes), and (c) choline C (δ 3.223 ppm, present in VLDL fraction).

In terms of the validity of the model, the Q^2 value for PLS-DA and OPLS-DA models were 0.56 and 0.52, respectively, which characterize good models to discriminate the pre-GD and control samples.

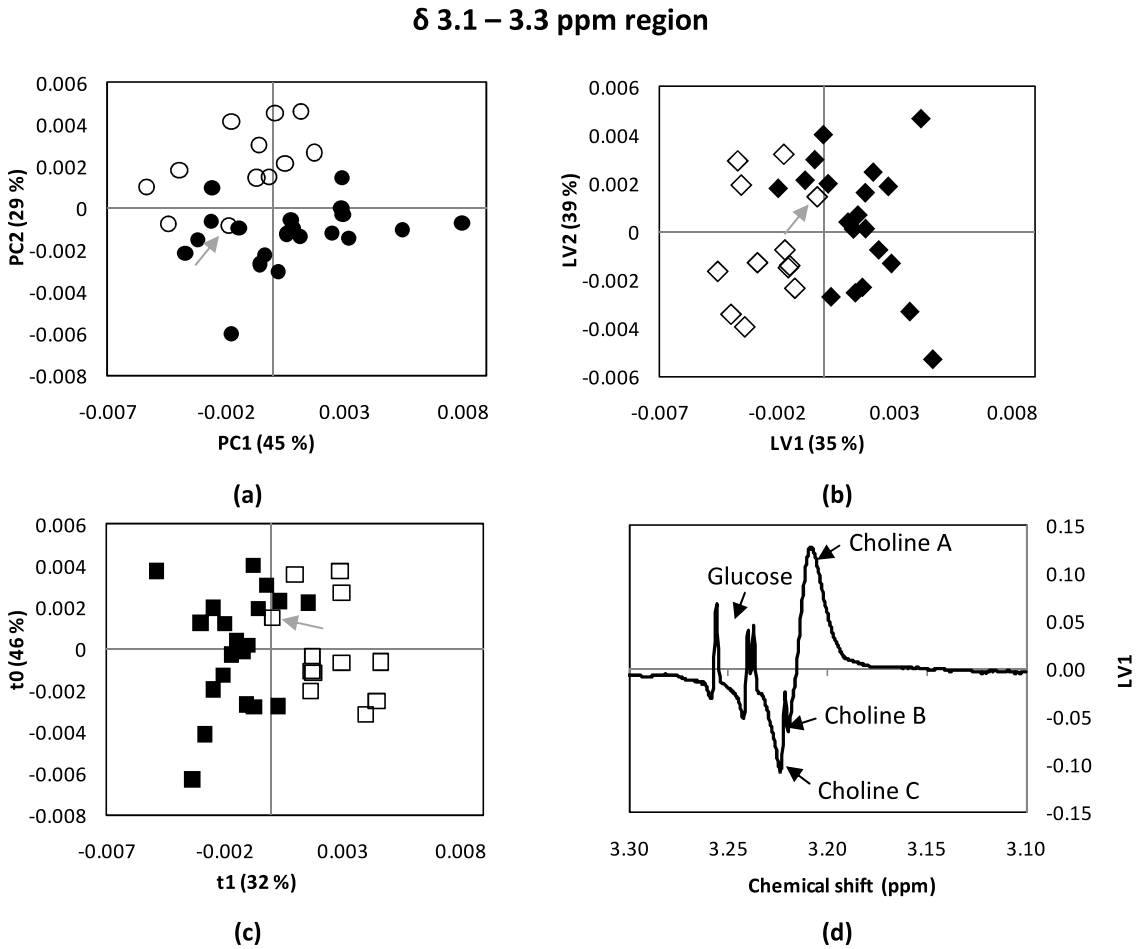


Figure 3.12 MVA of control (black symbols, n=20) and pre-GD (open symbols, n=12) samples in δ 3.1-3.3 region, as viewed by standard 1D ^{1}H NMR spectra of plasma. (a) PCA scores plot, PC=15; (b) PLS-DA scores plot, LV=2, R^2X=0.73, R^2Y=0.65, Q^2=0.56; (c) OPLS-DA scores plot, LV=1+1, R^2X=0.73, R^2Y=0.65, Q^2=0.52; and (d) OPLS-DA loading plot. PS 34109 sample is indicated by arrow.

The results obtained with the CPMG and diffusion-edited spectra were similar to those obtained with the standard 1D spectra, with choline compound A and C, and glucose to discriminate between control and pre-GD samples, as can be seen in Figure 3.13. The CPMG spectra provide a view mainly of the glucose resonance because choline resonances are attenuated due their shorter relaxation time. On the other hand, in the diffusion-edited spectra, the glucose resonance is strongly attenuated due to glucose higher diffusion coefficient and the choline resonances can be viewed more clearly.

The OPLS-DA model of diffusion-edited spectra discriminates better between the two sample classes (higher R^2X value) and in terms of Q^2, both diffusion-edited and standard 1D and diffusion-edited spectra models show higher values than with CPMG spectra.

In summary, MVA performed in the δ 3.1-3.3 ppm region gives good discrimination between control and pre-GD samples and suggests higher levels of choline compound A and glucose than in controls, and lower levels of choline compound C.

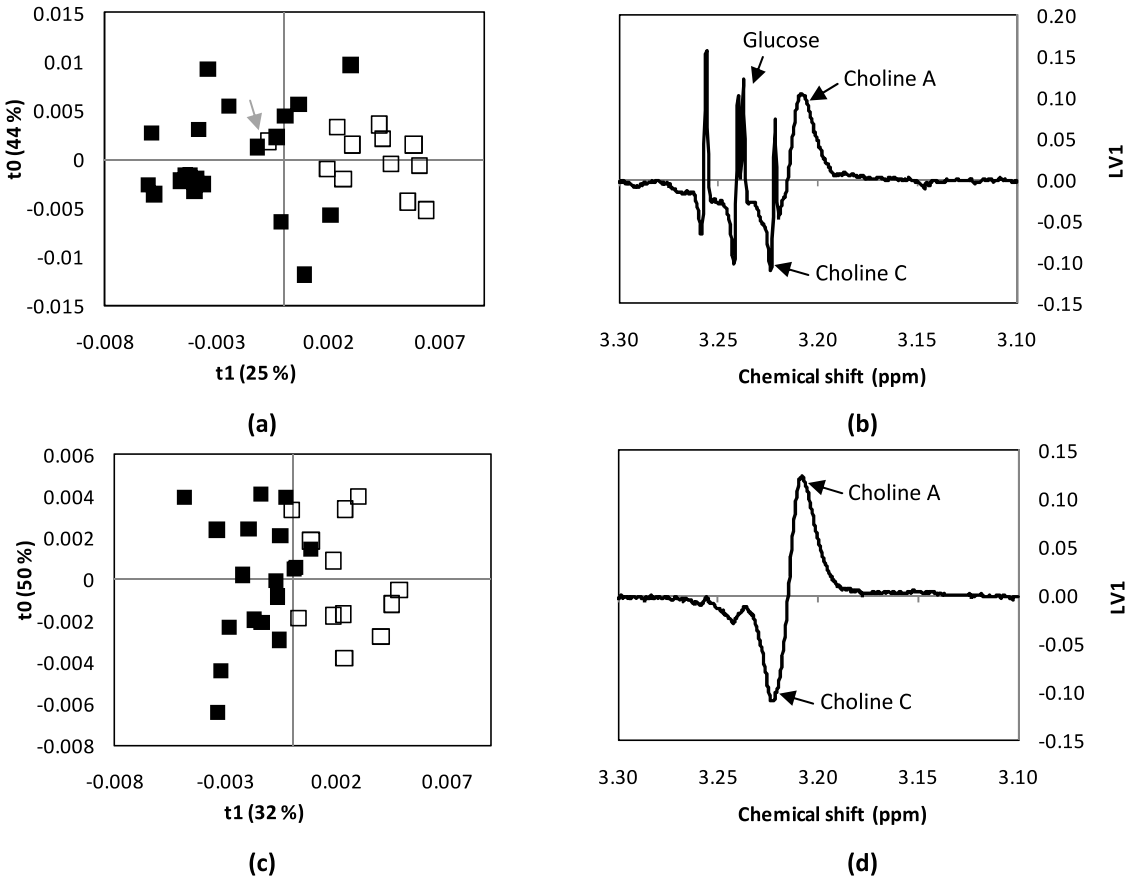


Figure 3.13 MVA of control (black symbols, n=20) and pre-GD (open symbols, n=12) samples in δ 3.1-3.3 region. (a) and (b) OPLS-DA scores plot and loading plot, respectively, as viewed by CPMG ^{1}H NMR spectra of plasma, LV=1+1, R^2X=0.69, R^2Y=0.60, Q^2=0.47; (c) and (d) OPLS-DA scores plot and loading plot, respectively, as viewed by diffusion-edited ^{1}H NMR spectra of plasma, LV=1+1, R^2X=0.82, R^2Y=0.62, Q^2=0.56.

The next step was the computing of the relevant peak areas to confirm the MVA findings. As shown above (Figure 3.11), the δ 3.1-3.3 ppm region is difficult to integrate because the resonances of β-glucose and choline are heavily overlapped. In order to overcome this, spectral deconvolution into the individual peaks previously identified by LC-NMR was carried out. The difference in glucose content between the two groups was verified through the doublet signal of α-glucose at δ 5.23 ppm, where there is less overlap.

The peak areas resulting from deconvolution were compared using boxplots (Figure 3.14). For both phospholipid choline compounds A (Figure 3.14a) and B (Figure 3.14b), there are a few outliers in both control and pre-GD groups and the variability within the pre-GD group is higher than in controls (which can be seen by the higher length of the box and the wiskers). Statistical tests were applied to the data, as described in experimental section. For choline compounds A and B, the results suggested that control and pre-GD group are significantly different. However, for choline C and glucose, the results suggested that control and pre-GD group are not significantly different. Therefore, the

choline B is significantly different between the two groups, the major difference is given by choline A (lower p-value) confirming the results obtained by MVA.

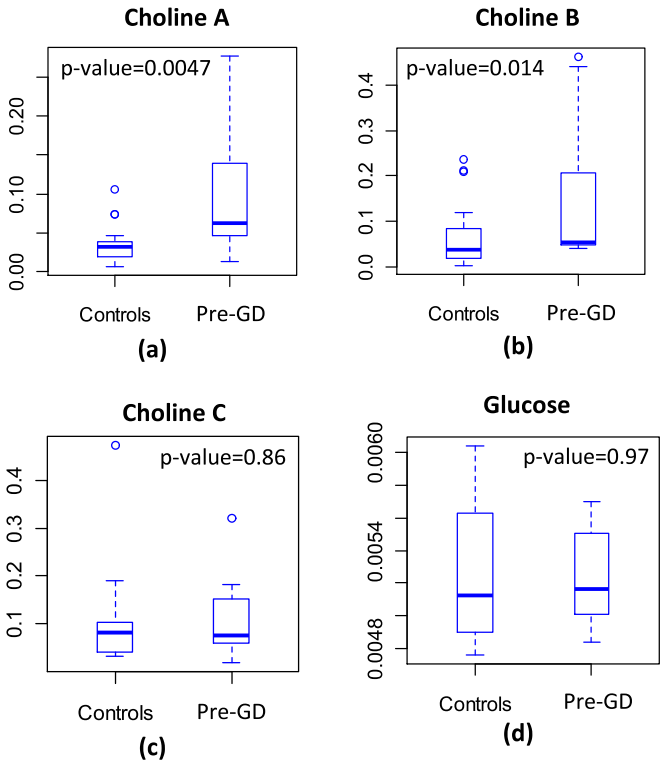


Figure 3.14 Boxplots of integral values obtained by deconvolution of peak at δ 3.217 (a), peak at δ 3.207 (b) and peak at δ 3.195 normalized to total spectral area.

Finally, the percentage of variance of each pre-GD sample relatively to controls was determined and the results are shown in Table 3.5. For choline A resonance, only two samples are inside the control confidence interval (samples with x in Table 3.5), the PS 34109 and PS 19108 which indicates that they have levels of choline A similar to those of control samples. In fact, the PS 34109 appears close to controls in the models discussed above. For choline B resonance, half of the number of the pre-GD samples are inside the control confidence interval, which means that they have integral values close to the average integrals of the control samples.

PS 65808 sample is outside of control confidence interval, however, shows lower levels of choline B than controls (negative percentage). Nevertheless, the results obtained by deconvolution and UVA confirm some of the results obtained by MVA, namely the indication of the high levels of choline B in pre-GD group, while the hypothesis of higher levels of glucose was discarded as significant.

In the future, another approach that can be used to confirm these results is the use of sensitive enzymatic measurements for choline compounds and glucose.

In the literature, there is evidence that choline is critical during fetal and neonatal life to ensure optimal brain and cognitive development. Thus, the high requirements of the fetus can be obtained from maternal stores of choline which is actively transported from mother to fetus across the placenta.[95] However, to our knowledge no relationship is known between GD and choline requirements/access. In type 2 diabetes mellitus, an elevation of two phosphoethanolamine species and a decrease in two lysophosphocholine species were found in plasma.[96]

Table 3.5 Variations of choline integrals in Pre-GD samples compared to controls[a]

Pre-GD samples	Choline A (%)	Choline B (%)
PS 6306	39.9	x
PS 8306	285	x
PS 19108	x	210
PS 19408	663.3	592.8
PS 59108	36.9	x
PS 60408	75.3	212.9
PS 65808	-61.3	133.7
PS 114008	287.9	x
PS 129208	247.6	559.2
PS 34109	x	x
PS 3709	328.3	x

[a] x indicates values inside the control confidence interval (95% confidence)

3.1.4. Fetal Malformations viewed by NMR

The aim of this work was also to search for biomarkers for FM during the second trimester of pregnancy from the ^{1}H NMR metabolic profile of mother's blood plasma. For these, ultrasound serves diagnostic requirements (although sometimes some ambiguity remains) but extra biochemical information may be useful for disease understanding. For this purpose, MVA was applied to the ^{1}H NMR spectra of women whose fetus had different types of malformations, either suspected at the time of amniocentesis. The types of malformations considered are presented in Table 3.6, and, as can be seen, the low number of samples of each type of malformation is a limiting factor in the study.

The strategy used in the study of FM was identical to that described for the pre-GD group. MVA was performed using the 10 samples in Table 3.6 for the FM group and 20 control samples from healthy pregnant women for the control group.

Table 3.6 Types of malformations considered in this study. w, weeks; d, days.

Samples	Gestational Age	Type of Malformation
PS 141705	17 w	Uro-genital
PS 143405	22 w	Central Nervous System
PS 706	21 w	Polimalformations
PS 1806	18 w	Choroid plexus cyst
PS 5906	21 w	Cystic Hygroma
PS 111108	18 w + 5 d	Polimalformations
PS 21509	23 w + 1 d	Abdominal
PS 26609	22 w	Clubfoot
PS 29309	24 w + 3 d	Uro-genital
PS 42109	25 w	Polimalformations
PS 42209	23 w + 3 d	Cardiac

The standard 1D and diffusion-edited spectra of plasma did not show any significant changes between the FM and the control groups, this indicates that no significant change occur in the macromolecule domain. However, the CPMG spectra showed trends between classes for full spectra (δ 0.5-10.0 ppm), and for four spectral regions δ 0.75-1.43, δ 1.4-3.1, δ 2.965-3.525, and δ 5.15-5.38 ppm.

Figure 3.15 shows the results obtained by MVA of full CPMG spectra with 0.01 ppm buckets and UV-scaled data. In the PCA scores plot (Figure 3.15a), the samples are uniformly distributed with no separation between the two groups. However, in the PLS-DA (Figure 3.15b) and OPLS-DA (Figure 3.15c) scores plots, there is a clear separation between the two groups. In the OPLS-DA loading plot (Figure 3.15d), the peaks coloured red and orange correspond to the metabolites with highest weight value in the discrimination between the FM and control groups. The loadings plot using full spectra has many peaks making assignment difficult, however, some assignments were achieved: acetate (δ 1.91 ppm) with positive LV1 loading, and unknown metabolite (δ 1.41 ppm), pyruvate (δ 2.36 ppm), mannose (δ 5.18 ppm), and histidine (δ 7.7ppm) with negative LV1 loadings. This means that FM samples in positive t_1 have higher levels of acetate and lower levels of unknown metabolite, pyruvate, mannose, and histidine.

In terms of OPLS-DA model validity, R^2X is low, while R^2Y is high, which together with low Q^2, suggests that a solid separation can not be considered.

The MVA using spectral regions can help to improve the results and confirm, or not, the hypothesis advanced by MVA of full CPMG ^{1}H NMR spectra.

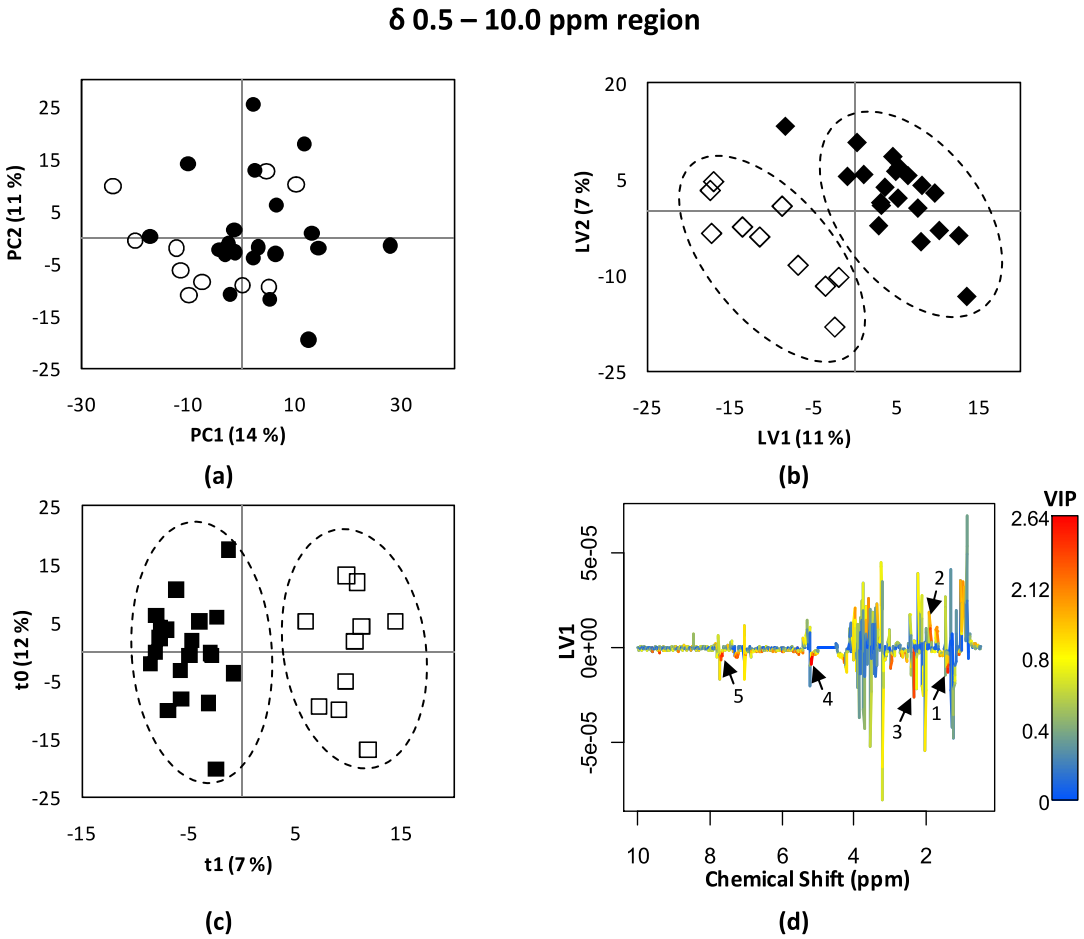


Figure 3.15 MVA of control (black symbols) and FM (open symbols) samples in δ 0.5 – 10.0 region, as viewed by the CPMG ^{1}H NMR spectra of plasma. (a) PCA scores plot, PC=5; (b) PLS-DA scores plot, LV=2, R^2X=0.19, R^2Y=0.90, Q^2=0.42; (c) OPLS-DA scores plot, LV=1+1, R^2X=0.19, R^2Y=0.90, Q^2=0.33; and (d) OPLS-DA loading plot. Key: 1, unknown metabolite; 2, acetate; 3, pyruvate; 4, mannose; and 5, histidine.

The spectral region at δ 0.75-1.43 ppm of CPMG ^{1}H NMR spectra with resonances of metabolites such as amino acids, lipoproteins and lactate showed a grouping trend. By visual inspection of the average CPMG spectra of controls (not shown) and FM, the spectral profiles are similar between the two groups. The MVA results are shown in Figure 3.16, with 0.01 ppm buckets and UV-scaled data.

A PCA model was built with eight PCs, and the samples are uniformly distributed in scores plot (Figure 3.16a) with no separation between the two groups. The PLS-DA (two LVs), and OPLS-DA (one orthogonal component) are also shown in Figure 3.16. In both scores plot, separation is clear between the two groups, except for one PS 706 sample. Interestingly, this sample belongs to a 17 years old woman at 21 weeks of gestation with a polymalformed fetus which led to early interruption of the pregnancy.

In terms of Q^2 value, PLS-DA is not a good model because a 0.046 of Q^2 value is very low. In the case of OPLS-DA, Q^2 value is 0.26, which is significantly better than for PLS-DA

but still too low. Hence, the results can only be considered as a slight trend and not as a solid separation between the two groups.

The LV1 loading plot of OPLS-DA is shown in Figure 3.16d. The metabolites with highest VIP are unknown metabolite (δ 1.41 ppm), threonine (δ 1.30 ppm), and from long chain $(CH_2)_n$ of fatty acids (δ 1.26 ppm), which are highest in FM and few controls; and valine (δ 1.03 ppm) and isoleucine (δ 0.95 ppm) which are lowest. The peak with high VIP indicated by number 4 in figure is an artefact due the chemical shift deviations in threonine at δ 1.30 ppm caused by pH differences among samples. Compared to results obtained for full spectra, this region allows the detection of long chain $(CH_2)_n$ of fatty acids, valine and isoleucine, and the confirmation of the unknown metabolite.

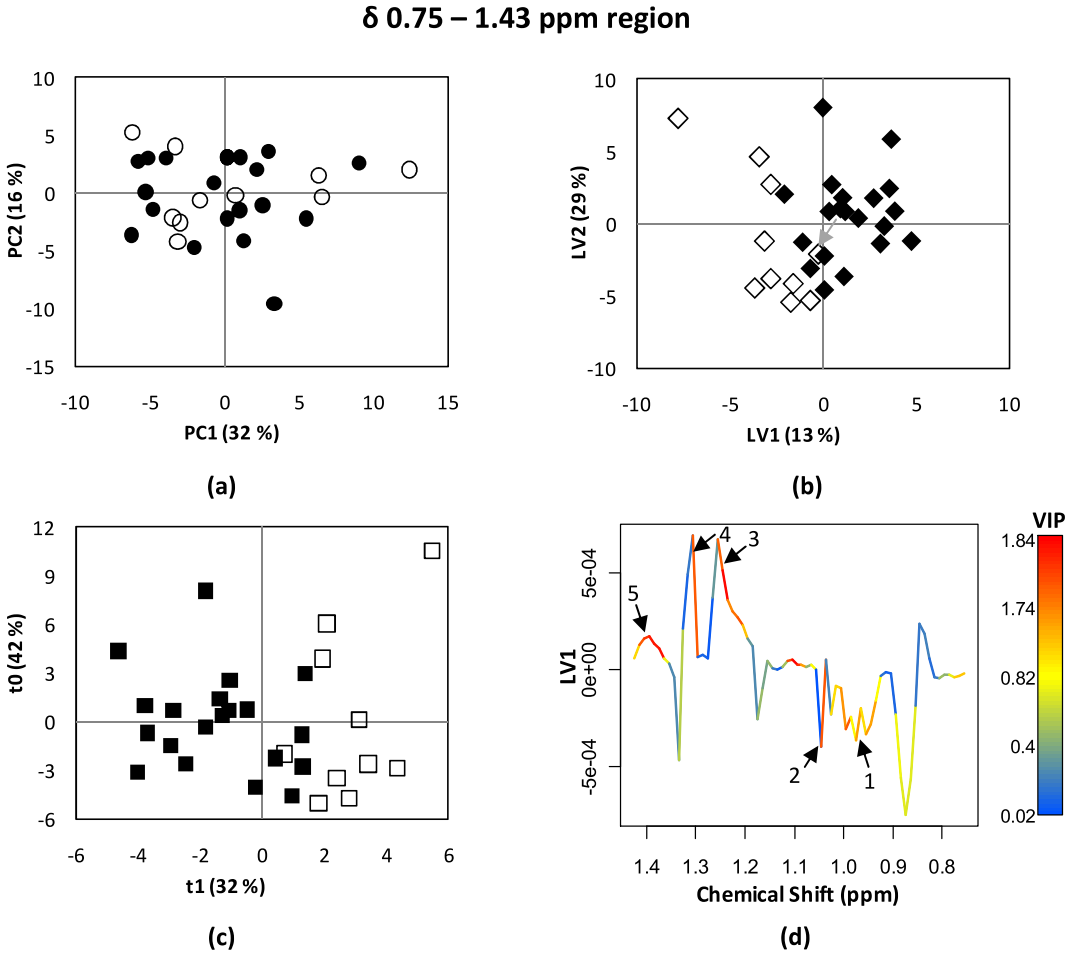


Figure 3.16 MVA of control (black symbols) and FM (open symbols) sample in δ 0.75-1.43 region, as viewed by the CPMG ^{1}H NMR spectra of plasma. (a) PCA scores plot, PC=8; (b) PLS-DA scores plot, LV=2, R^2X=0.42, R^2Y=0.58, Q^2=0.046; (c) OPLS-DA scores plot, LV=1+1, R^2X=0.42, R^2Y=58, Q^2=0.26; and (d) OPLS-DA loading plot. Key: 1, isoleucine; 2, valine; 3, long chain $(CH_2)_n$ of fatty acids; 4. threonine; 5. unknown metabolite.

In order to validate the results obtained by PLS-DA and OPLS-DA, UVA was performed and the resulting boxplots are shown in Figure 3.17. In the case of threonine, the doublet at δ 3.60 ppm was used in UVA, because the peak at δ 1.30 ppm is overlapped by the doublet of lactate at δ 1.32 ppm. UVA confirms that there are significantly differences in the unknown metabolite (Figure 3.17a), threonine (Figure 3.17b), valine (Figure 3.17d) and isoleucine (Figure 3.17e). For the long chain $(CH_2)_n$ fatty acids of lipoproteins (Figure 3.17c), UVA did not show significantly differences between the two groups. Therefore, most of the hypothesis advanced by MVA was confirmed by UVA, except for long chain $(CH_2)_n$ fatty acids of lipoproteins. Table 3.7 provides a summary of the results for the δ 0.75-1.43 ppm region of CPMG ^{1}H NMR spectra.

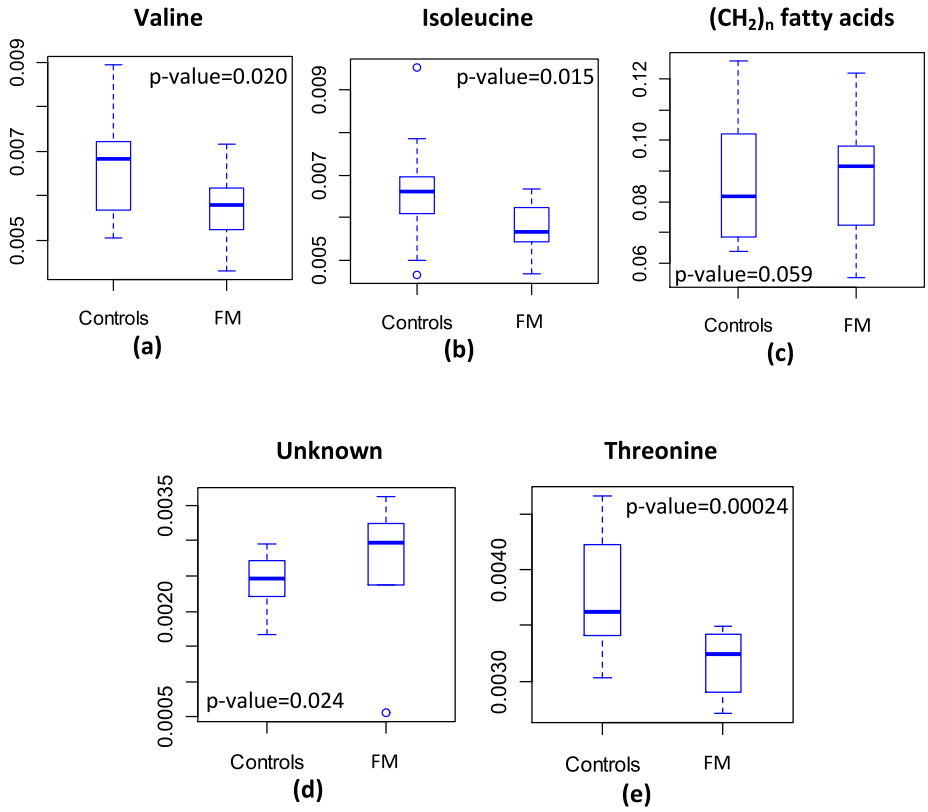


Figure 3.17 Boxplots of normalized integral values of metabolites with highest VIP obtained by OPLS-DA analysis of δ 0.75-1.43 ppm region of CPMG ^{1}H NMR spectra of control and FM group.

Table 3.7 Summary of the hypothesis advanced by MVA and the results obtained by UVA of δ 0.75-1.43 ppm region of CPMG ^{1}H NMR spectra of control and FM group.

Analysis of δ 0.75-1.43 ppm region	
MVA	**UVA**
Threonine	✓ Threonine
Unknown	✓Unknown
Long chain fatty acids	× Long chain fatty acids
Valine	✓ Valine
Isoleucine	✓ Isoleucine

Another interesting result was found for the δ 1.4-3.1 ppm region of CPMG ^{1}H NMR spectra (Figure 3.18) with 0.01 ppm buckets and Pareto-scaled data. This spectral region is dominated by resonances from low molecular metabolites such as alanine, acetate, acetoacetate, pyruvate, creative and creatinine. Comparing the average CPMG spectra of controls and FM, some relative intensity differences can be observed for acetate, acetoacetate and pyruvate, as indicate by arrows in Figure 3.18.

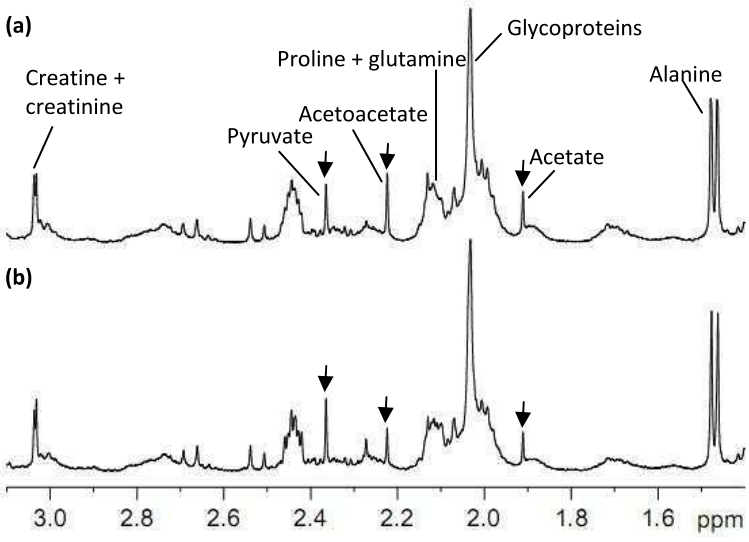


Figure 3.18 CPMG ^{1}H NMR average spectra of plasma in δ 1.4-3.1 region of (a) control and (b) FM groups.

In the PCA scores plot (not shown) no sample groupings were visible. The results obtained by PLS-DA and OPLS-DA were similar and Figure 3.19 shows the latter. Indeed, a separation can be viewed between FM in positive t1 and controls in negative t1, however there are two outlier malformation samples (PS 706 and PS1806). The PS 706 sample appeared close to controls already in the δ 0.75-1.43 ppm region discussed above. The PS 1806 sample belongs to a mother whose fetus had choroid plexus cyst but no further clinical information is available. The Q^2 value obtained was 0.26 which is quite low. The OPLS-DA loading plot (Figure 3.19b) indentifies the metabolites with highest importance: alanine (δ 1.46 ppm), lysine and arginine (δ 1.72 ppm), acetate (δ 1.91 ppm), glycoprotein (δ 2.03 ppm), acetoacetate (δ 2.22 ppm), and pyruvate (δ 2.36 ppm). The region indicated by number 5 in figure is again an artefact due the chemical shift variations of glutamine and methionine at δ 2.10 ppm. Thus, pyruvate and glycoprotein seem to be higher in FM than in controls (positive LV1), and acetoacetate, acetate, lysine, arginine and alanine are lower in FM than in controls (negative LV1). This spectral region allows the confirmation of acetate and pyruvate suggested by the full CPMG spectra, and the detection of glycoprotein, acetoacetate, lysine, arginine and alanine.

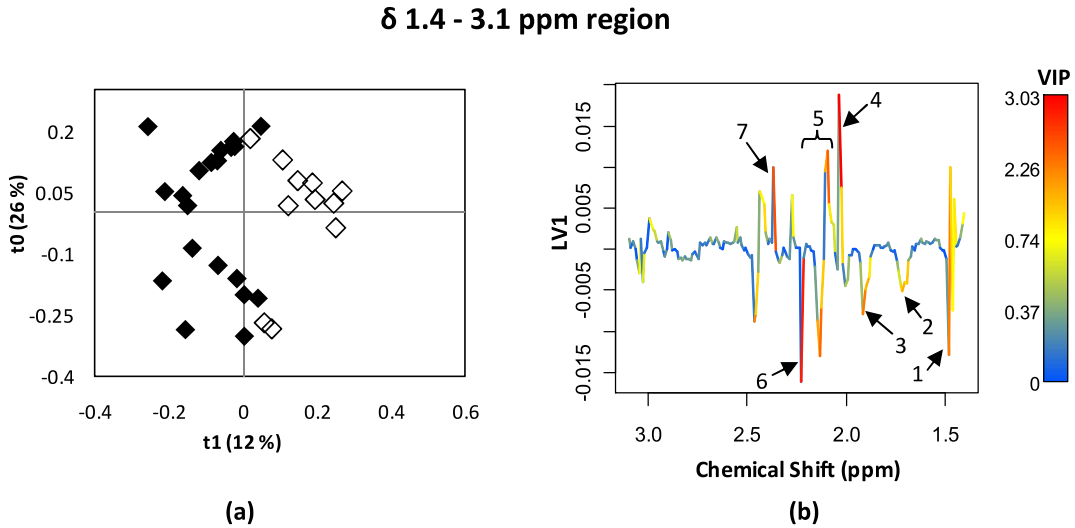


Figure 3.19 MVA of control (black symbols) and FM (open symbols) samples in δ 1.4-3.1 ppm region, as viewed by CPMG ^{1}H NMR spectra. (a) OPLS-DA scores plot LV=1+1, R^2X=0.38, R^2Y=0.65, Q^2=0.26; and (b) OPLS-DA loading plot. Key: 1, alanine; 2, lysine and arginine; 3, acetate; 4, glycoprotein; 5, proline and glutamine; 6, acetoacetate; and 7, pyruvate.

Then, UVA was performed on the metabolites with highest VIP, and the boxplots are shown in Figure 3.20. For lysine and arginine (Figure 3.20b), acetate (Figure 3.20c), glycoprotein (Figure 3.20d) and pyruvate (Figure 3.20f), significant differences are found

between controls and FM, and this confirms the results obtained by MVA. However, for alanine (Figure 3.20a) and acetoacetate (Figure 3.20c), there are no significant differences between controls and FM by UVA. A summary of the results for the δ 1.4-3.1 ppm region of CPMG ^{1}H NMR spectra is shown in Table 3.8.

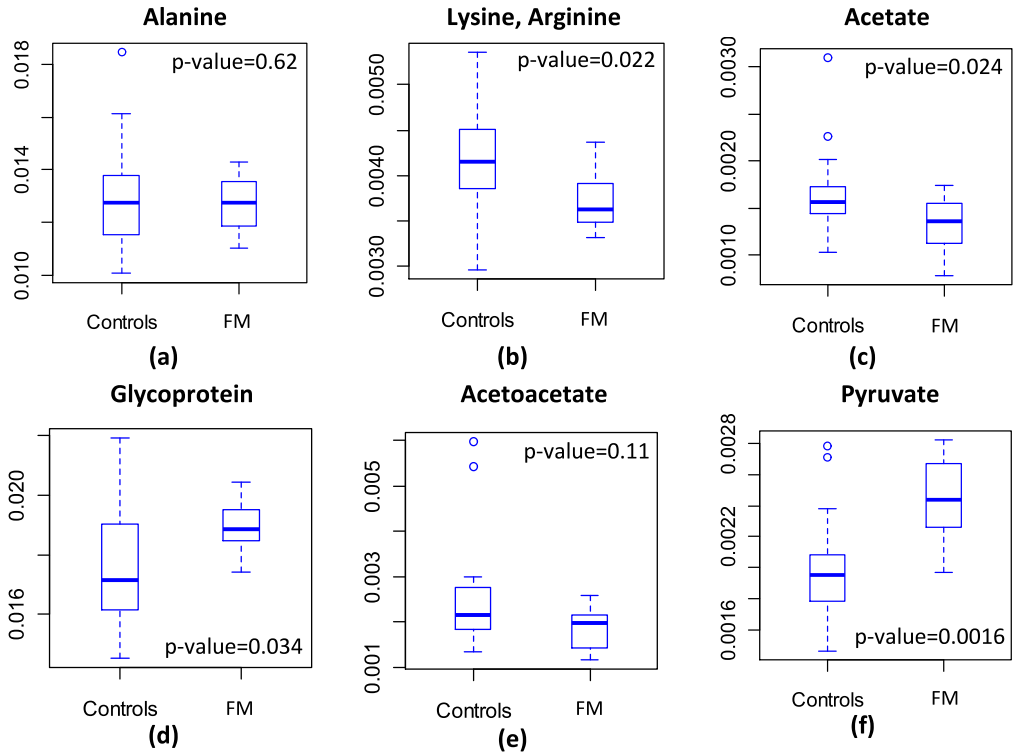


Figure 3.20 Boxplots of normalized integral values of metabolites with highest VIP obtained by OPLS-DA analysis of δ 1.4-3.1 ppm region of CPMG ^{1}H NMR spectra of control and FM group.

Table 3.8 Summary of the hypothesis advanced by MVA and the results obtained by UVA of δ 1.4-3.1 ppm region of CPMG ^{1}H NMR spectra of control and FM groups.

Analysis of δ 1.4-3.1 ppm region	
MVA	**UVA**
Alanine	× Alanine
Lysine, arginine	✓ Lysine, arginine
Acetate	✓ Acetate
Glycoprotein	✓ Glycoprotein
Acetoacetate	× Acetoacetate
Pyruvate	✓ Pyruvate

In addition to the MVA studies described above, iPLS-DA also revealed good discrimination between the two groups identifying the δ 2.965-3.525 ppm region of

CPMG ^{1}H NMR spectra as more relevant. However, the CPMG average spectra do not show visual differences between the controls and FM samples. The iPLS-DA model was built with five LVs, and in the scores plot (Figure 3.21a) a grouping of FM samples can be seen of FM samples in LV1 positive, except for two samples, PS 1806 and PS 141705. Although, some controls are also in positive LV1, it is a good model due the high Q^2 value (0.65).

The iPLS-DA loading plot (Figure 3.21b) shows that the signals contributing for discrimination are at δ 3.215 and δ 3.225 ppm, which according to the names defined in previous section (3.1.3) for pre-GD, are the choline compound B (present in HDL, LDL and VLDL particles) and choline compound C (present in VLDL). Thus, the iPLS-DA results suggest that there is considerably more choline B in FM than in controls. Since, PLS-DA and OPLS-DA models did not show any trend between the two groups in δ 2.925-3.525 ppm region, iPLS-DA complements the information obtained by both PLS-DA and OPLS-DA.

To confirm the results obtained by iPLS-DA, deconvolution was performed with the information obtained by LC-NMR. The standard 1D ^{1}H NMR spectra were used instead of the CPMG ^{1}H NMR spectra, because in the latter there is considerable attenuation of choline resonances. The boxplot obtained for choline B confirms that the FM group has higher levels of choline B compared to controls, however, the choline C is not significantly different between the two groups (Figure 3.22).

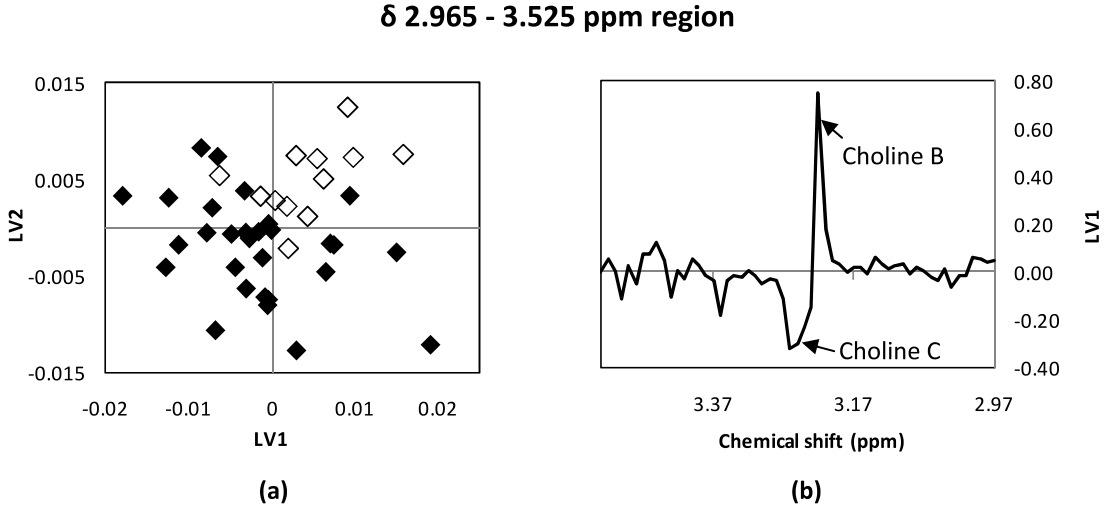


Figure 3.21 iPLS-DA scores plot (a) and loading plot (b) of controls (black symbols, n=29) and FM (open symbols, n=12) samples, as viewed by the CPMG ^{1}H NMR spectra of plasma between δ 2.965-3.525 using 0.01 ppm buckets and no-scaling. LV=5, Q^2=0.65.

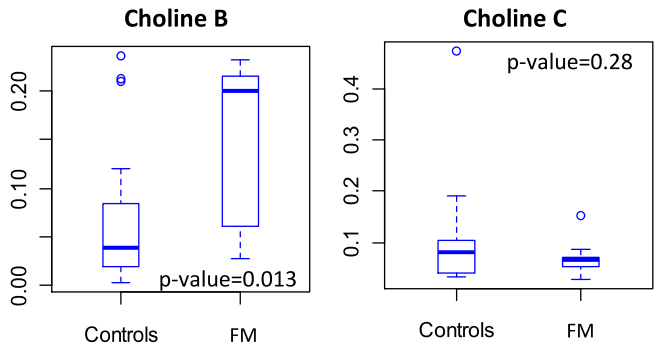


Figure 3.22 Boxplot of choline C normalized integral values obtained by deconvolution in δ 3.1-3.3 ppm region of standard 1D ^{1}H NMR spectra of control and FM groups.

In the δ 5.15-5.38 ppm region of the CPMG ^{1}H NMR spectra, without bucketing and UV-scaled data, another trend was found. In this region there are a few resonances, the unsaturated bonds of lipoprotein-lipids, α-glucose and mannose. It was noted that the CPMG average spectra show no apparent differences in the spectral profile of control and FM groups.

In the PCA model (not shown), the scores plot did not show groupings between the two groups. The results obtained by PLS-DA and OPLS-DA were similar and the results of OPLS-DA can be seen in Figure 3.23. The scores plot (Figure 3.23a) shows all FM samples grouped in positive t_1, together with a few controls. Interpretation of the OPLS-DA loading plot (Figure 3.23b) identifies a single metabolite with highest VIP. This is mannose (δ 5.18 ppm), which is higher in FM than in controls as can be confirmed by the boxplot obtained shown in Figure 3.24.

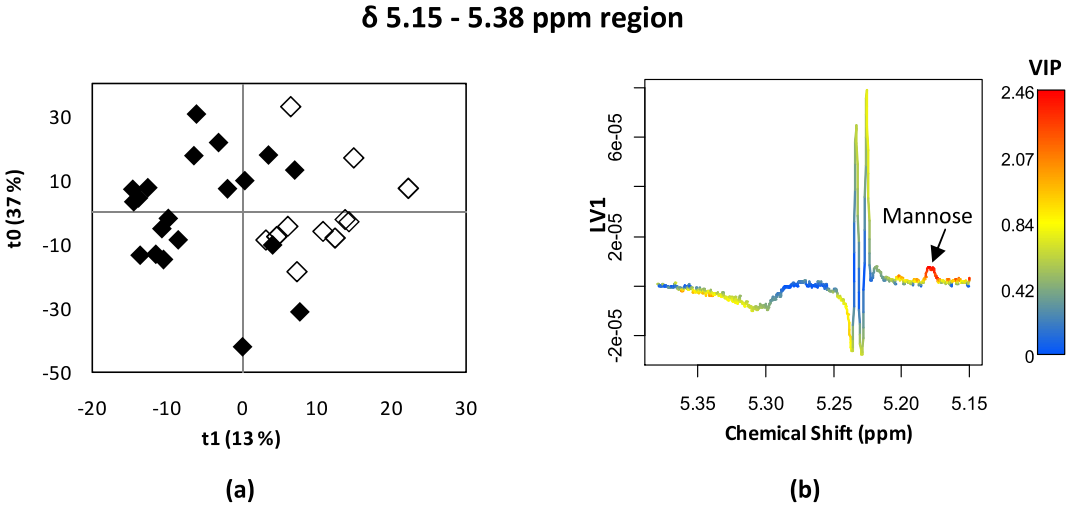


Figure 3.23 MVA of control (black symbols) and FM (open symbols) samples in δ 5.15-5.38 ppm region, as viewed by CPMG ^{1}H NMR spectra of plasma. (a) OPLS-DA scores plot, LV=1+1, R^2X=0.50, R^2Y=0.58, Q^2=0.20; and (b) OPLS-DA loading plot.

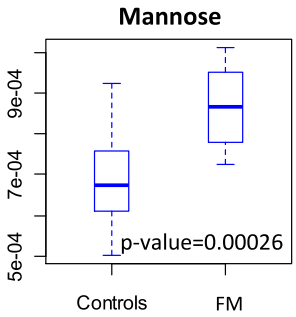


Figure 3.24 Boxplot of normalized integral value of mannose for control and FM groups.

Finnally, recally the MVA of full CPMG ^{1}H NMR spectra (Figure 3.15), it is noted that analysis of smaller spectral regions allowed the further confirmation of four metabolites (unknown metabolite, acetate, pyruvate and mannose) and the detection of seven metabolites (valine, isoleucine, glycoprotein, lysine, arginine, threonine and choline B). Table 3.9 provides a summary of all the results obtained by MVA and confirmed by UVA for FM group. The results suggest that five amino acids (treonine, valine, isoleucine, lysine and arginine) are present in lower levels in blood plasma of mothers with FM when compared to controls, as well as acetate. Mannose, pyruvate, and unknown metabolite show higher levels in FM group. The major difference observed between the two groups is for choline B levels which show differences at around 200%. In addition, the differences in glycoprotein levels can not be considered significant because the levels of most FM samples are within the control confidence interval.

Table 3.9 Variations of important metabolites in FM samples compared to controls[a]

Metabolite	Val	Ile	Unknown	Glyco-protein	Pyruvate	Acetate	Lys, Arg	Mannose	Thr	Choline B
PS21509	-13.1	-19.5	49.4	15.2	19.9	-10.7	-9.9	x	-8.3	211.8
PS111108	-12.3	-13.2	26.4	x	33.9	-49.0	-19.2	17.9	-7.3	237
PS26609	-15.7	x	33.6	10.2	14.6	-27.3	-15.2	36.4	-26.5	x
PS29309	-6.9	x	37.2	10.2	37.5	x	x	44.8	-9.1	200.3
PS141705	-28.3	-28.8	x	x	x	-22.5	-14.8	17.5	-22.8	158.2
PS42109	7.9	x	20.3	x	13.3	-12.9	x	38.8	-13.7	247.1
PS42209	-20.8	-14.7	24.7	x	26.3	-40.8	-15.8	47.4	-13.9	201.2
PS143405	-35	-17.3	x	8.2	41.6	-11.2	-19.9	13.5	-27.8	x
PS1806	-6.8	-14.9	11.0	x	24.6	13.1	-7.1	34.6	-16.2	-59.1
PS706	x	-9.9	-77.0	9.9	x	x	x	11.0	-11.4	222.9

[a] x indicates values inside the control confidence interval (95% confidence). Abbreviations: Val, valine; Ile, Isoleucine; Lys, lysine; Arg, arginine; Thr, Threonine.

Biochemical interpretation of these changes is not an easy task. Some suggestions are identified below but more objective interpretation would require more specific studies, such as enzymatic activity measurements.

As mentioned above, choline is a very important metabolite during fetal development being used to ensure optimal brain development. Metabolically, choline is known to be a precursor of betaine, and both metabolites are important for adequate liver function, cellular replication, and detoxification reactions.[95]

The source of blood mannose is unknown, at this stage. This metabolite is required for glycoprotein and glycolipids biosynthesis. A study in embryonic kidneys of mouse has suggested that mannose may cause energy depletion by cellular ATP consumption and this alters the synthesis of glycosylated proteins, such as proteoglycans, resulting in renal dysmorphogenesis.[97-98]

Amino acids are very important nutrients to fetus and their depletion may be caused by spare to fetus for support growth, since there is an active transport of amino acids from mother to fetus by the placenta.[99]

Pyruvate is an intermediate compound in the metabolism of carbohydrates, proteins and fatty acids. Acetate is involved in ethanol degradation, fatty acid biosynthesis and pyruvate metabolism.[34]

3.2. Study of Prenatal Disorders by FTIR Spectroscopy

In this study, the FTIR spectroscopy was applied to investigate its use as a high throughput profiling technique to determine if it is possible to discriminate plasma of women with prenatal disorders, such as pre-GD and FM, and healthy women.

The first step at this stage was the optimization of the experimental procedure in order to achieve short acquisition times and good reproducibility. As described in the experimental section, sample drying at room temperature (24^{o}C) was attempted, as well as drying with a stream of nitrogen. The results obtained for a healthy pregnant woman (PS 9906) at 16 weeks of gestation are shown in Figure 3.25. As can be seen, in the FTIR spectrum obtained directly, i. e., without drying (0 min), mainly characteristic bands of water (3600-3200 cm^{-1} and 1650 cm^{-1}) and CO_2 (2400-2200 cm^{-1}) can be visualized. The blood plasma is 91.5% water, hence, in order to detect the minor components one solution would be the spectral subtraction but it often originates artefacts in the spectrum, therefore it was found preferable to dry the sample.

Figure 3.25a shows the spectra acquired over time for drying at room temperature. Comparing the spectra acquired at 20, 30 and 40 min, there are no differences between them (all superimposed in figure) which suggest that the plasma sample was dry after 20 min. The drying with a stream of nitrogen (Figure 3.25b) show strong differences between all spectra acquired over time during 45 min. It was expected that the drying with stream of nitrogen would be more rapid, however, reproducible were not achieved. One explanation may be the uncontrolled sweeping of the sample by the stream which produces variability in the amount of sample placed on the crystal. Thus, the method chosen for drying the blood plasma samples was at room temperature.

The FTIR-ATR spectra of a dried blood plasma sample can be assigned as follows: the 4000 to 3100 cm^{-1} region contains only the broad infrared bands of water from the O-H stretching and H-O-H bending vibrations; from 3100 to 2800 cm^{-1}, fatty acid CH, CH_2 and CH_3 symmetric and asymmetric stretching vibrations are visible reflecting the presence of triglycerides, cholesterol esters, long chain and unsaturated fatty acids and phospholipids; the 1750 to 1500 cm^{-1} region comprises C=O stretching of amide I and N-H bending vibrations of amide II in amino acids and proteins. The 1500 to 1250 cm^{-1} region accommodates contributions from CH_2 and CH_3 symmetric and asymmetric bending vibrations of fatty acids, phospholipids and triglycerides and COO^{-} stretching vibrations of amino acids. Finally, the 1200-900 cm^{-1} region shows absorbances from C-O stretching vibrations of glucose, lactate and glycerol.

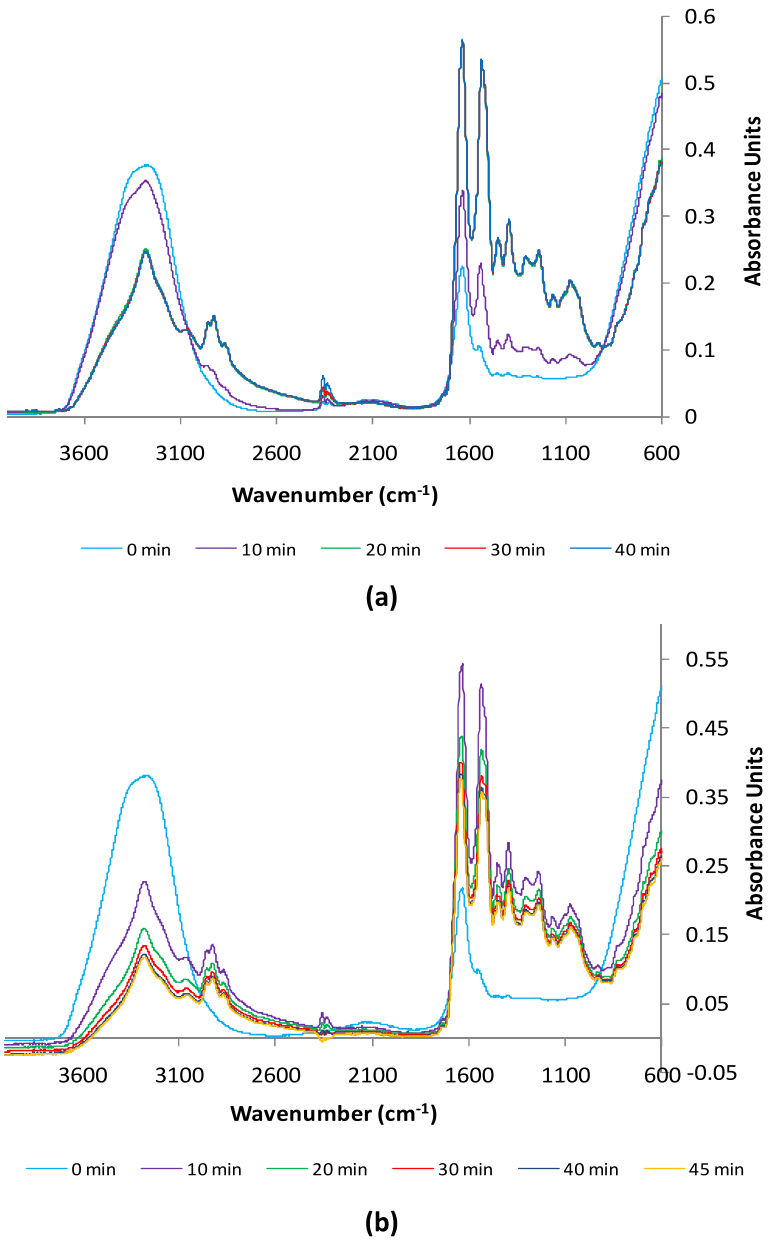


Figure 3.25 FTIR-ATR spectra of drying of a healthy pregnant woman sample (PS 9906) at 16 weeks of gestation. (a) Spectra acquired with sample drying at room temperature (24 °C); (b) Spectra acquired with sample drying using a stream of nitrogen.

PCA was applied to the spectra acquired at 20, 30 and 40 min, for all samples, in order to confirm the best time from which the samples may be considered as completely dry. The resulting PCA model (Figure 3.26) shows that the spectra acquired at 30 and 40 min are grouping, while the spectra of 20 min is more distant from them. This suggests that from 30 min onwards the plasma samples were effectively dried.

Figure 3.26b shows the PCA model built using three aliquots of each of five different samples acquired in different days, in order to test the reproducibility between different days of measurement. As can be seen, in three of the samples there is one aliquot more distant relative to the others, as indicated by dashed circles in figure. The variability

between these aliquots is explained by variations in room relative humidity (RH) which, unfortunately, was difficult to control. In fact, it is observed that grouping samples were acquired at 33 % RH, the PS 133505 and PS 11406 samples were acquired at 30% RH, and PS 10306 at 43% RH. Hence, the humidity seen to be determinant factor for reproducibility and should be maintained for all samples acquired.

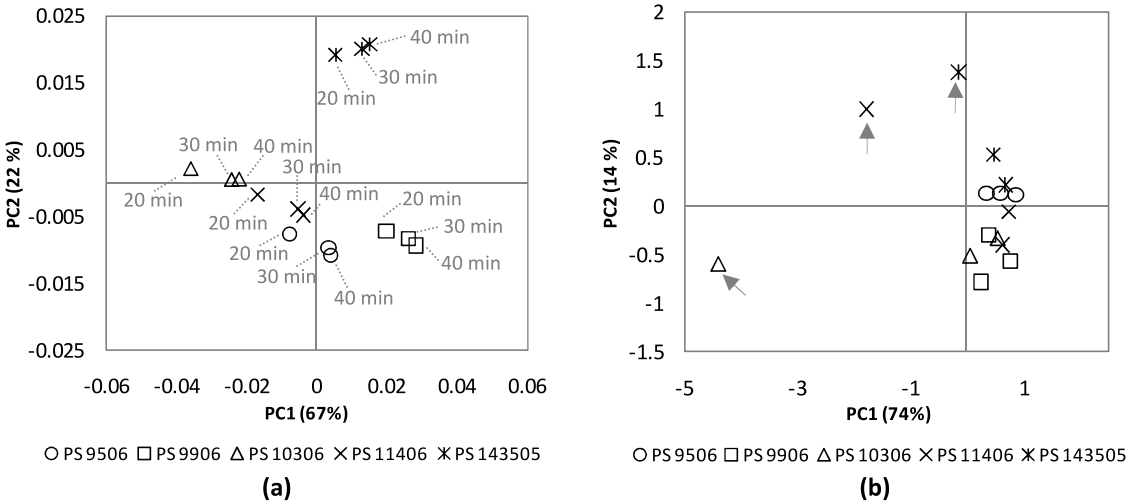


Figure 3.26 (a) PCA model of spectra acquired at 20, 30 and 40 min for five control samples; (b) PCA model of spectra of three aliquots acquired in different days for five control samples. The arrows indicate outlier aliquots.

In Figure 3.27, a FTIR spectrum of plasma of a healthy pregnant woman obtained in this work and the spectrum of blood plasma of an adult man found in literature[74] can be seen. The two spectra are in general very similar, however, slightly differences in 1800-900 cm^{-1} region are visible mainly for N-H bending vibrations of amide II.

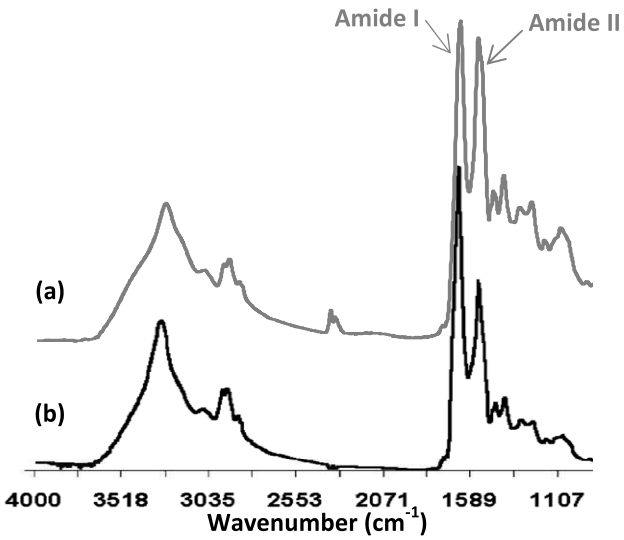


Figure 3.27 Comparison of FTIR spectra of (a) plasma sample of a healthy pregnant woman at 16 weeks of gestation (PS 9906, gray) acquired in this work with drying at room temperature and (b) plasma sample of an adult man acquired with drying vacuum (black) [74].

3.2.1. Gestational Diabetes Viewed by FTIR

As described in the experimental section, MVA was performed in all absorbance spectra, the first derivative absorbance spectra and the second derivative absorbance spectra. The results presented here are only those obtained from MVA performed to the second derivative absorbance spectra, since studies on the normal absorbance spectra and the first derivative absorbance spectra did not show good discrimination between the classes studied. The resulting PLS-DA models obtained for the second derivative full spectra (4000-600cm^{-1}), 3110-2800cm^{-1}, and 1200-900cm^{-1} regions (Figure 3.28) clearly show groupings between the two classes. However, the explained variance in LV1 and LV2 is considerably low (around 10%) mainly for 3110-2800cm^{-1} and 1200-900cm^{-1} regions.

Comparing the R^2X, R^2Y and Q^2 values given in Table 3.10, the models 3110-2800 cm^{-1} and 1200-900 cm^{-1} regions have very low R^2X and high R^2Y values. For full spectra, the R^2X is better than for spectral regions and the Q^2 value is very high, which characterize a valid model. However, the interpretation of the PLS-DA loadings plots is not easy. In PLS-DA loading plot of full spectra (Figure 3.28b), the wavenumbers with highest VIP were at 668 cm^{-1}, in positive LV2, and at 665 cm^{-1}, in negative LV2, in the fingerprint region. In PLS-DA loading plot of 3110-2800 cm^{-1} region (Figure 3.28d), the wavenumber with highest VIP was at 2847 cm^{-1} in negative LV2, thus suggesting that pre-GD samples have more long chain fatty acids and phospholipids than control samples. Indeed, this result agrees with the results obtained by MVA performed to ^{1}H NMR spectra of control and pre-GD groups, which shows high levels of choline-containing phospholipids in lipoprotein particles for pre-GD group. Finally, the PLS-DA loading plot of the 1200-900 cm^{-1} region (Figure 3.28f) shows wavenumbers with highest VIP at 1085 and 1031 cm^{-1}, in positive LV1, and 1049 cm^{-1}, in negative LV1, which are in the fingerprint region.

In summary, the discrimination between the pre-GD and control groups is mostly attributed to the fingerprint region where the detailed interpretation of the infrared bands is very difficult. The results obtained in the 3110-2800 cm^{-1} region suggests that pre-GD samples have high levels of long chain fatty acids and phospholipids, which are in agreement with the results obtained by NMR.

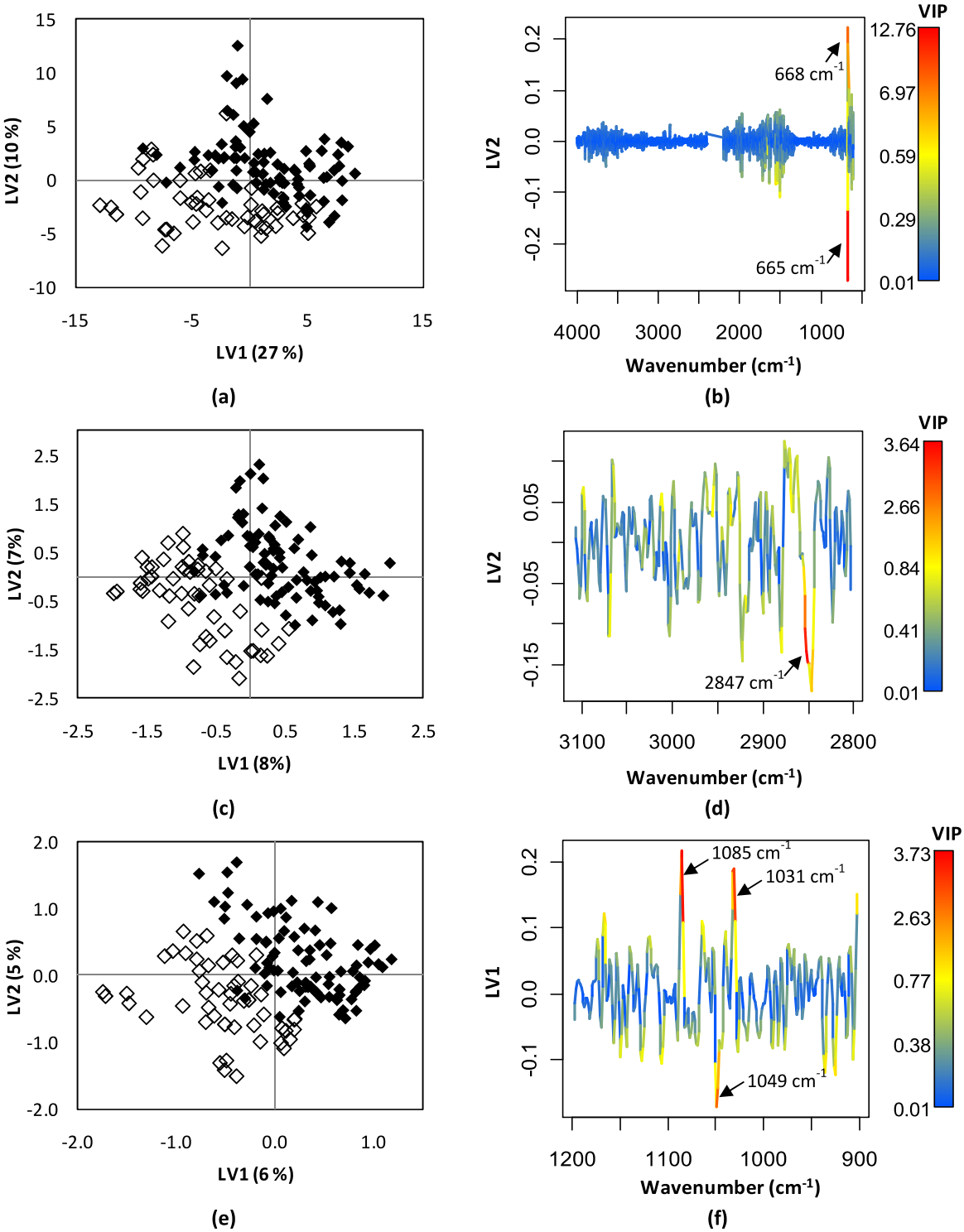


Figure 3.28 PLS-DA of control (black symbols, n=20) and pre-GD (open symbols, n=12) samples of the data derived from the second derivative of the FTIR spectra. (a), (c) and (e) PLS-DA scores plot of full spectra (4000-600 cm^{-1}), 3110-2800 cm^{-1} and 1200-900 cm^{-1} regions, respectively. (b), (d) and (f) PLS-DA loadings plots full spectra (4000-600 cm^{-1}), 3110-2800 cm^{-1} and 1200-900 cm^{-1} regions, respectively.

Table 3.10 Results obtained from PLS-DA of second derivative FTIR spectra performed to control and pre-GD group

Region	PLS-DA			
	LV	R^2X	R^2Y	Q^2
4000-600 cm^{-1}	5	0.46	0.94	0.81
3110-2800 cm^{-1}	3	0.18	0.77	0.49
1200-900 cm^{-1}	2	0.11	0.64	0.38

3.2.2. Fetal Malformations Viewed by FTIR

The MVA for FM group was performed in the same way as for pre-GD group. The results presented here, like for pre-GD, are only from the MVA performed to the second derivative absorbance spectra because the results using the normal and the first derivative absorbance spectra not showed good discriminations.

For FM, the spectral regions with good discrimination were the same as for pre-GD, the 3110-2800 cm^{-1} and 1200-900 cm^{-1} regions. In PCA model of both regions, the samples were uniformly distributed with no separation between the FM and control groups. On the other hand, PLS-DA was successful to discriminate the two classes and the results are represented in Figure 3.29.

The number of LVs, R^2X, R^2Y and Q^2 values is summarized in Table 3.11. The explained variance in LV1 and LV2 is again considerably low, mainly for 3110-2800 cm^{-1} and 1200-900 cm^{-1} regions. Comparing the R^2X, R^2Y and Q^2 values with the scores plot for each spectral region, the full spectra gave the best discrimination between the two groups.

In PLS-DA loading plot of full spectra (Figure 3.29b), the wavenumbers with highest VIP are at 681 cm^{-1}, in positive LV2, and at 1650 cm^{-1} and 677 cm^{-1} in negative LV2. The wavenumbers at 681 and 677 cm^{-1} are in the fingerprint region. The wavenumber at 1650 cm^{-1} suggests differences in proteins.

For 3110-2800 cm^{-1} and 1200-900 cm^{-1} regions, the PLS-DA loadings plots (Figure 3.29d and f) can not be interpreted because the wavenumbers with highest VIP are uniformly distributed between the positive and negative LV1. This fact makes impossible to interpret which variable is responsible for the variation between the two groups.

In summary, as well as for pre-GD group, the absorbance bands in fingerprint region are responsible for grouping trends. However, the results obtained for full FTIR spectra also suggest that FM group has higher protein content than in controls.

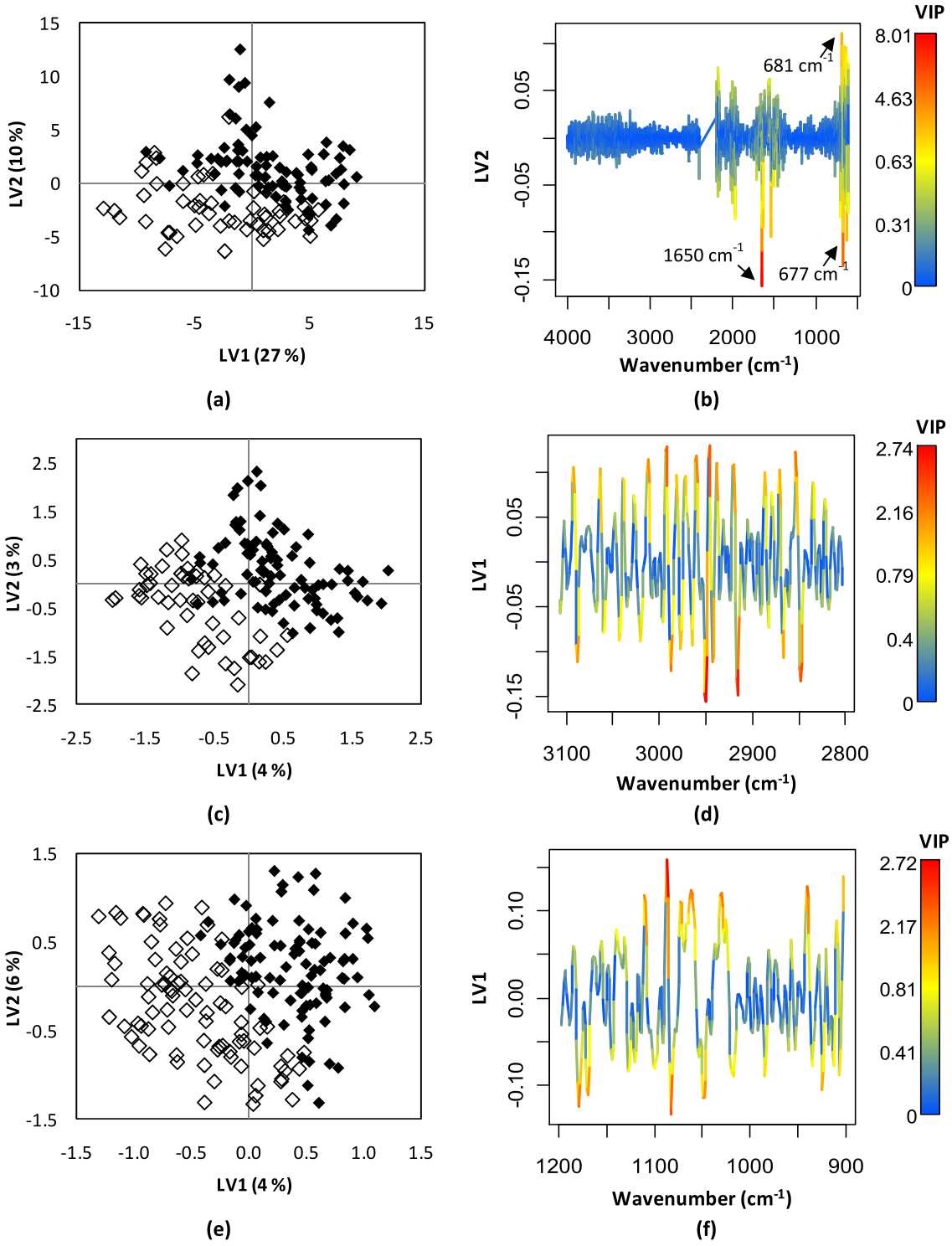


Figure 3.29 PLS-DA of control (black symbols, n=20) and FM (open symbols, n=17) samples of the data derived from the second derivative of the FTIR spectra. (a), (c) and (e) PLS-DA scores plot of full spectra (4000-600 cm^{-1}), 3110-2800 cm^{-1} and 1200-900 cm^{-1} regions, respectively. (b), (d) and (f) PLS-DA loadings plots of full spectra (4000-600 cm^{-1}), 3110-2800 cm^{-1} and 1200-900 cm^{-1} regions, respectively.

Table 3.11 Results obtained from PLS-DA of second derivative FTIR spectra performed to control and FM groups.

	PLS-DA			
Region	**LV**	**R^2X**	**R^2Y**	**Q^2**
4000-600 cm^{-1}	6	0.66	0.94	0.77
3110-2800 cm^{-1}	2	0.07	0.68	0.35
1200-900 cm^{-1}	3	0.15	0.75	0.50

In this work the assignment of the ^{1}H NMR spectra of blood plasma was carried out using several 1D and 2D NMR experiments, as well as, the information found in literature allowing the identification of 12 amino acids, 8 organic acids, glucose, mannose, choline, albumin, glycoproteins and lipoproteins. However, the assignment of the main lipoprotein classes was not easily achieved directly, due to their similar chemical shift profiles, and LC-NMR spectroscopy was then applied with the aim to separate the major lipoprotein classes and confirm their respective chemical shift assignment. The chemical shift information thus obtained allowed deconvolution of the spectra of whole blood to be carried out and determination of the relative percentages of the lipoproteins and choline-containing compounds present. The results showed that a selected blood plasma sample contained 24% of HDL, 25% of LDL, 27% of VLDL and 24% of saturated fatty acids. The specific assignment of the different choline-containing compounds was not possible at this stage, but it was observed that the HDL fraction showed higher percentages of choline compounds than the LDL and VLDL fractions. In the future, it is important to determine the precision of these measurements and extend them to routine analysis of whole blood plasma lipoproteins.

MVA was carried out on the NMR spectra of whole blood plasma to study the metabolic changes present in maternal blood plasma collected at 15-24 weeks of gestation for women who later were diagnosed with GD (named here as pre-GD subjects) or whose fetuses were affected by several malformations. In spite of the relatively number of samples available, these preliminary results show the potential of NMR-metabonomics to provide insight into the biochemical profile of these prenatal diseases. When considering pre-GD, significant grouping was noted by PCA, PLS-DA and OPLS-DA in the three types of ^{1}H NMR experiments acquired. The results suggest higher levels of a choline compound resonating at δ 3.209 ppm in HDL and another resonating at δ 3.223 ppm in VLDL fraction. To our knowledge, no other reports of choline-compound alterations have been reported in relation to pre-GD or diagnosed GD subjects. The FM group reveals less pronounced alterations in maternal blood plasma, relatively to controls, and only slight groupings were noted by PLS-DA and OPLS-DA in CPMG ^{1}H NMR spectra. The metabolite changes suggested were lower levels of amino acids (valine, isoleucine, threonine, lysine and arginine) and acetate, and higher levels of pyruvate, mannose, and choline compound at δ 3.215 ppm present in the three major lipoprotein classes (as viewed by LC-NMR).

In a final stage of the work, FTIR spectroscopy in tandem with MVA was also carried out to investigate the ability of this technique to distinguish between control and prenatal diseases blood plasma. Interestingly, significant separation between control and diseased groups were observed by PLS-DA for both pre-GD and FM. However, these were only achieved using the second-derivative absorbance spectra which means that the alterations in the spectra are of small magnitude. For both pre-GD and FM groups, grouping trends were visible when considering the full spectra (4000-600 cm^{-1}) and the

3110-2800 cm^{-1} and 1200-900 cm^{-1} regions of the second-derivative spectra. The wavenumbers with highest importance in this discrimination were mainly found in the fingerprint region and, thus, their specific assignment is difficult. However, in the case of the pre-GD group, the 3110-2800 cm^{-1} region suggests that pre-GD samples have higher levels of long chain fatty acids and phospholipids, in agreement with the results obtained by NMR, whereas for FM group , higher protein contents.

In the future, the confirmation of these results using more specific studies, such as sensitive enzymatic measurements, and the further understanding of these changes may allow maternal plasma metabolites to be used as biomarkers/predictors of the diseases.

Appendix 1. General Clinical Information of Plasma Samples Studied

Samples were coded according to: PS is the abbreviation for “plasma sanguíneo”, the next digits are the sample code given in prenatal diagnosis service, and the last two digits are the year of sample collection.

Table 0.1 General information of control plasma samples used in this study. W, weeks; d, days.

Sample	Gestational Age	Maternal Age
PS 142205	17 w	35
PS 143505	16 w	30
PS 143905	17 w	36
PS 144005	17 w	35
PS 2006	17 w	31
PS 3006	18 w	41
PS 3206	16 w	38
PS 4706	17 w	35
PS 6606	17 w	36
PS 6706	17 w	36
PS 7206	17 w	31
PS 8006	16 w	38
PS 8106	18 w	29
PS 8206	18 w	32
PS 8806	17 w	36
PS 9906	16 w	35
PS 10306	18 w	35
PS 11306	17 w	39
PS 11406	15 w	36
PS 35608	17 w + 5 d	36
PS 35808	16 w + 3 d	37
PS 123108	17 w + 1 d	34
PS 125808	17 w + 1 d	39
PS 129608	16 w + 2 d	36
PS 130208	16 w + 3 d	35
PS 135708	16 s + 6 d	36
PS 137608	17 w	17
PS 4009	18 w + 4 d	37
PS 7309	23 w + 5 d	39
PS 33609	-	30

Table 0.2 General information of pre-GD and FM plasma samples used in this study. W, weeks; d, days.

	Sample	Gestational Age	Maternal Age	Relevant clinical information
Pre-GD	PS 6306	17 w	39	
	PS 8306	17 w	35	
	PS 19108	17 w + 1 d	35	
	PS 19408	17 w + 5 d	40	
	PS 59108	-	31	
	PS 60408	16 w + 5 d	32	
	PS 65808*	17 w + 1 d	37	
	PS 114008	16 w + 3 d	30	Risk of tromboembolism
	PS 129208	16 w + 5 d	39	Malformed fetus
	PS 3709	17 w + 2 d	40	
	PS 34109	21 w	35	
	PS sc109	16 w	35	
FM	PS 141705	17 w	26	Uro-genital
	PS 143405	22 w	32	Central Nervous System
	PS 706**	21 w	17	Polimalformations: tetramelia, bilateral cleft lip and cleft palate
	PS 1806	18 w	32	Choroid Plexus Cyst
	PS 5906**	21 w	21	Cystic Hygroma
	PS 111108*	18 w + 5 d	27	Polimalformations: bilateral clubfeet, choroid plexus cyst, pielactasia (left, 8mm)
	PS 21509	23 w + 1 d	27	Abdominal
	PS 26609	22 w	31	Clubfoot
	PS 29309	24 w + 3 d	31	Uro-genital
	PS 42109	25 w	27	Polimalformations: cerebellar hypoplasia, macrocephaly, cleft lip and palate, chromosomopathy 46 XY, ish del(3) (qter)
	PS 42209	23 w + 3 d	24	Cardiac, Trissomy 21
	PS 47109	17 w +4 d	19	Clubfoot
	PS 53309	35 w +5 d	24	Central Nervous System
	PS 123009	22 w +1 d	32	Cardiac
	PS SLA109	32 w +5 d	-	Central Nervous System
	PS SLA209	23 w	-	Cardiac
	PS SLA309	17 w +6 d	-	Hyper Echogenic Intestine
	PS 40810	23 w +5 d	-	Central Nervous System

* woman with medication, **interruption of the pregnancy

Appendix 2. ^{1}H and ^{13}C NMR Assignments of Human Blood Plasma

Table 0.3 Resonance assignments with chemical shifts and spin-spin coupling patterns of metabolites identified in 750 MHz and ^{1}H-^{13}-C NMR spectra of normal human blood plasma [70].

^{1}H shift (δ)	Multiplicity	Molecule	Assignment	Observed	^{13}C shift (δ)
0.66	m	cholesterol	C18 (in HDL)	1D, HMQC	12.6
0.70	m	cholesterol	C18 (in VLDL)	HMQC	
0.84	m	cholesterol	C26 and C27	HMQC	23.3
0.84	t	lipid (mainly LDL)	CH_3 (CH_2)n	1D, JRES, COSY, HMQC	14.7
0.87	t	lipid (mainly VLDL)	$CH_3CH_2CH_2C$=	1D, JRES, COSY	
0.91		cholesterol	C21	HMQC	19.4
0.93	m	lipid	CH_3CH_2	COSY	
0.93	t	isoleucine	δ-CH_3	1D, JRES, COSY	
0.95	d	leucine	δ-CH_3	1D, JRES, COSY	
0.97	d	leucine	δ-CH_3	1D, JRES, COSY	
0.97	d	valine	CH_3	1D, JRES, COSY, HMQC	19.6
1.00	d	isoleucine	β- CH_3	1D, JRES, COSY, HMQC	14.6
1.02	d	valine	CH_3	1D, JRES, COSY	
1.13	d	isobutyrate	CH_3	CPMG, JRES	
1.20	d	3-hydroxybutyrate	γ CH_3	1D, CPMG, JRES	
1.22	m	lipid	$CH_3CH_2CH_2$	HMQC	32.7
1.25	m	lipid (mainly LDL)	$(CH_2)_n$	1D, CPMG, JRES, COSY, HMQC	30.6
1.26	m	lipid	$CH_3CH_2(CH_2)_n$	COSY, HMQC	23.2
1.28	m	isoleucine	half γ-CH_2	COSY	
1.29	m	lipid (mainly VLDL)	$CH_2CH_2CH_2CO$	1D, CPMG, JRES, COSY	
1.30	m	lipid	CH_2	COSY, HMQC	19.7
1.31	d	fucose	CH_3	CPMG, JRES, COSY	
1.32	d	threonine	γ-CH_3	JRES, COSY	
1.32	m	lipid	$CH_2CH_2CH_2CO$	COSY	
1.33	d	lactate	CH_3	JRES, COSY, HMQC	20.9
1.46	d	alanine	CH_3	1D, CPMG, JRES, COSY, HMQC	16.8
1.47	m	isoleucine	half γ -CH_2	JRES, COSY	
1.48	m	lysine	γ -CH_2	CPMG, COSY, HMQC	
1.57	m	lipid (mainly VLDL)	CH_2CH_2CO	JRES, COSY, HMQC	25.6
1.57	m	citrulline	γ -CH_2	1D, COSY	
1.68	m	arginine	γ -CH_2	1D, CPMG, COSY	
1.69	m	lysine	δ-CH_2	1D, COSY	27.4
1.69	m	lipid	CH_2CH_2C=C	1D, JRES, HMQC	40.7

Table 0.3 (Continued)

^{1}H shift (δ)	Multiplicity	Molecule	Assignment	Observed	^{13}C shift (δ)
1.71	m	leucine	β-CH_2, γCH	COSY, HMQC	
1.86	m	citrulline	β -CH_2	COSY	
1.91	m	lysine	β -CH_2	JRES, COSY, HMQC	30.3
1.91	m	arginine	β -CH_2	COSY	
1.91	s	acetate	CH_3	JRES, CPMG	
1.96	m	isoleucine	β-CH	COSY	
1.97	m	lipid	CH_2C=C	1D, COSY	
1.99	m	proline	γ-CH_2	COSY	
2.00	m	lipid	CH_2C=C	COSY, HMQC	27.8
2.00	m	lipid	CH_2C=C	COSY	
2.00	m	glutamate	Half β-CH_2	JRES, COSY	
~2.04	s	glycoprotein[b] (acetyls)	$NHCOCH_3$	1D, JRES, HMQC	23.0
2.05	m	proline	half β-CH_2	COSY	
2.08	m	glutamine	half β -CH_2	1D, JRES, COSY	
2.09	m	glutamine	half β -CH_2	1D, JRES, COSY	
2.13	s	methionine	S-CH_3	CPMG, JRES	
2.14	m	glutamate	half β -CH_2	1D, JRES, COSY, HMQC	30.1
2.22	s	acetoacetate	CH_3	CPMG, JRES	
2.23	m	lipid	CH_2CO	COSY, HMQC	34.6
2.24	spt of d	valine	β-CH	COSY	
2.31	m	3-hydroxybutyrate	half α-CH_2	CPMG, JRES, COSY	
2.36	m	glutamate	half γ-CH_2	COSY, HMQC	
2.36	s	pyruvate	CH_2	1D, CPMG, JRES, CPMG	
2.36	m	proline	half β-CH_2	COSY	34.5
2.38	m	3-hydroxybutyrate	half α-CH_2	JRES, COSY	
2.39	ABX	U1[c]	CH_2	COSY	
2.41	m	glutamine	half γ-CH_2	CPMG, JRES, COSY, HMQC	31.9
2.47	t	2-oxoglutarate	γ-CH_2	JRES	
2.52	d	citrate	half CH_2	JRES, COSY	
2.54	s	methylamine	CH_3	CPMG, JRES	
2.54	ABX	U1[c]	CH_2	COSY	
2.68	dd	aspartate	half β-CH_2	JRES	
2.69	d	citrate	half CH_2	JRES, COSY	
2.69	m	lipid	C-CCH_2C=C	COSY	
2.71	s	dimethylamine	CH_3	CPMG, JRES	
2.71	m	lipid	C=CCH_2C=C	1D, COSY	
2.72	m	lipid	C=CCH_2C=C	1D, COSY, HMQC	26.2
2.81	dd	aspartate	half β-CH_2	JRES	

Table 0.3 (Continued)

^{1}H shift (δ)	Multiplicity	Molecule	Assignment	Observed	^{13}C shift (δ)
2.83	s	trimethylamine	CH_3	1D, CPMG, JRES	
2.84	dd	asparagine	half β-CH_2	JRES	
2.89	t	albumin lysyl	ϵ-CH_2	1D, COSY, HMQC	40.3
2.94	dd	asparagine	half β-CH_2	JRES	
2.96	t	albumin lysyl	ϵ-CH_2	1D, COSY, HMQC	40.3
3.01	t	albumin lysyld	ϵ-CH_2	1D, JRES, COSY, HMQC	40.3
3.04	s	creatine	CH_3	CPMG, JRES	
3.05	s	creatinine	CH_3	1D, CPMG, JRES	
3.06	dd	tyrosine	half β- CH_2	JRES, COSY	
3.12	dd	phenylalanine	half β- CH_2	JRES	
3.14	dd	histidine	half β- CH_2	JRES	
3.15	t	citrulline	γ- CH_2	JRES	
3.16	dd	tyrosine	half β- CH_2	JRES, COSY	
3.21	s	choline	$N(CH_3)_3$	JRES, HMQC	55.0
3.24	t	arginine	δ- CH_2	COSY, HMQC	41.3
3.24	dd	β-glucose	H2	1D, JRES, COSY, HMQC	75.1
3.25	dd	histidine	half β-CH_2	JRES	
3.25	t	taurine	CH_2NH	JRES	
3.26	s	TMAO	CH3	JRES	
3.26	t	U2[c]	ABX	COSY	
3.26	dd	phenylalanine	half β-CH_2	COSY	
3.28	t	myo-inositol	H5	JRES	
3.34	m	proline	half δ-CH_2	COSY	
3.40	t	β-glucose	H4	1D, JRES, COSY, HMQC	70.6
3.41	t	taurine	CH_2SO3	JRES	
3.42	t	α-glucose	H4	JRES, COSY, HMQC	70.6
3.45	m	proline	half δ-CH_2	COSY	
3.47	ddd	β-glucose	H5	JRES, COSY, HMQC	76.7
3.48	dd	threonine	α-CH	JRES, COSY	
3.48	t	β-glucose	H3	JRES, COSY, HMQC	76.7
3.54	dd	α-glucose	H2	JRES, COSY, HMQC	72.3
3.54	dd	U2[c]	CH	COSY, HMQC	
3.54	s	glycine	CH_2	CPMG, JRES	
3.56	dd	myo-inositol	H1, H3	JRES	
3.56	dd	glycerol	half CH_2	1D, JRES, COSY	63.5
3.57	d	valine	α-CH	JRES, COSY, HMQC	64.2
3.60	d	threonine	α-CH	JRES	
3.63	dd	myo-inositol	H4, H6	JRES	
3.64	dd	glycerol	half CH_2	1D, JRES, COSY	63.5
3.66	m	choline (lipid)	NCH_2	COSY, HMQC	66.7
3.68	t	glutamine	α-CH_2	JRES, COSY, HMQC	55.4
3.69	dd	leucine	α-CH	JRES, COSY, HMQC	55.1

Table 0.3 (Continued)

^{1}H shift (δ)	Multiplicity	Molecule	Assignment	Observed	^{13}C shift (δ)
3.70	m	citrulline	α-CH	COSY	
3.71	t	α-glucose	H3	JRES, COSY, HMQC	73.6
3.72	dd	α-glucose	half CH_2-C6	1RES. COSY. HMQC	61.6
3.75	m	U1[c]	α-CH	COSY	
3.76	m	α-glucose	half CH_2-C6	JRES, COSY, HMQC	61.4
3.76	q	alanine	α-CH	IRES. COSY	
3.83	ddd	α-glucose	H5	JRES, COSY, HMQC	72.3
3.84	m	α-glucose	half CH_2-C6	JRES, COSY, HMQC	61.4
3.87	m	glycerol	C2-H	JRES, COSY	72.6
3.90	dd	β-glucose	half CH_2-C6	JRES, COSY, HMQC	61.6
3.93	s	creatine	CH_2	JRES	
3.94	dd	tyrosine	α-CH	JRES	
3.97		phenylalanine	α-CH	JRES, COSY	
3.98	dd	histidine	α-CH	JRES	
3.98	ABX	U2[c]	CH	COSY, HMQC	
4.05	s	creatinine	CH_2	1D, JRES	
4.06	m	glyceryl of lipids	CH_2OCOR	HMQC	62.5
4.06	t	myo-inositol	H2	JRES	
4.11	q	lactate	CH	1D, JRES, COSY, HMQC	69.2
4.12	m	proline	α-CH	JRES. COSY	
4.13	m	3-hydroxybutyrate	β-CH	1D, JRES, COSY	
4.24	m	threonine	β-CH	1D, JRES, COSY	
4.25	m	glyceryl of lipids	CH_2OCOR	HMQC	62.5
4.29	m	choline (lipid)	OCH_2	HMQC	60.2
4.53	d	β-galactose	H1	JRES	
4.64	d	β-glucose	H1	1D, CPMG, JRES, COSY, HMQC	
5.20	m	glyceryl of lipids	CHOCOR	COSY	
5.23	d	α-glucose	H1	1D, CPMG, JRES, COSY, HMQC	92.9
5.23	m	unsaturated lipid	$CH=CHCH_2CH=CH$	1D, COSY	128.6
5.26	m	unsaturated lipid	$CH=CHCH_2CH=CH$	1D, COSY, HMQC	128.6
5.27	m	unsaturated lipid	$=CHCH_2CH_2$	1D, COSY	
5.29	m	unsaturated lipid	$CH=CHCH_2CH-CH$	1D, COSY	128.6
5.31	m	unsaturated lipid	$=CHCH_2C_{H2}$	1D, COSY, HMQC	130.1
5.33	m	unsaturated lipid	$=CHCH_2CH_2$	1D, COSY	
6.87	s	tyrosine	H3, H5	1D, CPMG, HMQC	116.7
7.01	s	3-methylhistidine	H4	1D, CPMG, JRES	
7.02	s	histidine	H4	1D, CPMG, JRES	
7.05	s	1-methylhistidine	H4	1D, CPMG, JRES	
7.17	m	tyrosine	H2, H6	CPMG, JRES, COSY	

Table 11 (Continued)

^{1}H shift (δ)	Multiplicity	Molecule	Assignment	Observed	^{13}C shift (δ)
7.33	m	phenylalanine	H2, H6	1D, CPMG, COSY	
7.38	m	phenylalanine	H4	1D, CPMG, COSY	
7.43	m	phenylalanine	H3, H5	1D, CPMG, COSY	
7.61	s	methylhistidine	H2	1D, CPMG, JRES	
7.73	s	histidine	H2	1D, CPMG	
7.77	s	1-methylhistidine	H2	CPMG, JRES	
8.45	s	formate	CH	1D, CPMG	

[a] Abbreviations and Key: s, singlet; d, doublet; t, triplet; q, quartet; spt, septet; m, complex multiplet; dd, doublet of doublets; ddd, doublet of doublets of doublets. Chemical shifts all referenced to H-1 and C-1 of α-glucose at 5.233 for ^{1}H and at 92.9 ppm for ^{13}C. [b] Mainly α1-acid glycoprotein. [c] U1 and U2 refer to unidentified metabolites. [d] Indicates overlap of free lysine with albumin lysyl and α1-acid glycoprotein signals

References

1. Hod M., Jovanovic L., Di Renzo G. C., Leiva A. and L. O., *Textbook of Diabetes and Pregnancy*. 2nd ed. 2008, London: Informa Healthcare.
2. Moore, K.L. and T.V.N. Persaud, *The Developing Human: Clinically Oriented Embryology*. 2008, Philadelphia, PA: Elsevier Saunders.
3. Aziz, R. and T. Mahboob, *Pre-eclampsia and lipid profile.* Pakistan Journal of Medical Sciences, 2007. **23**(5): p. 751-754.
4. Catarino, C., I. Rebelo, L. Belo, P. Rocha-Pereira, S. Rocha, E.B. Castro, B. Patricio, A. Quintanilha and A. Santos-Silva, *Fetal lipoprotein changes in pre-eclampsia.* Acta Obstetricia et Gynecologica Scandinavica, 2008. **87**(6): p. 628-634.
5. Fraser V. J., Burd L. and Liebson E., *Diseases and Disorders*. 2007: Marshall Cavendish.
6. McGrowder, D., K. Grant, R. Irving, L. Gordon, T. Crawford, R. Alexander-Lindo and Y.T.P. Fraser, *Lipid Profile and Clinical Characteristics of Women with Gestational Diabetes Mellitus and Preeclampsia.* Journal of Medical Biochemistry, 2009. **28**(2): p. 72-81.
7. Wiznitzer, A., A. Mayer, V. Novack, E. Sheiner, H. Gilutz, A. Malhotra and L. Novack, *Association of lipid levels during gestation with preeclampsia and gestational diabetes mellitus: a population-based study.* American Journal of Obstetrics and Gynecology, 2009. **201**(5): p. 8.
8. Fischbach F. T., D., M. B., *A Manual of Laboratory and Diagnostic Tests*. 8th ed. 2009: Wolters Kluwer Health | Lippincott Williams & Wilkins.
9. Orshan, S.A., *Maternity, newborn, and women's health nursing: comprehensive care across the life span*. 2008: Lippincott Williams & Wilskins.
10. Carr D. B., G.S., *Gestational Diabetes: Detection, Management, and Implications.* Clinical Diabetes, January-February 1998. **16**(1).
11. Keeling, J.W. and T.Y. Khong, *Fetal and Neonatal Pathology*. 4th ed. 2007, London: Spinger.
12. Metzler, D.E., *Biochemistry: The Chemical Reactions of Living Cells*. 2nd ed. Vol. 1 and 2. 2003: Elsevier.
13. Scanlon, V.C. and T. Sanders, *Essentials of Anatomy and Physiology*. 5th ed. 2007: F. A. Davis Company.
14. Koolman, J. and K.H. Roehm, *Color Atlas of Biochemistry*. 2nd ed. 2005: Thieme.
15. Fischbach F. and Dunning M. B., *A Manual of Laboratory and Diagnostic Tests*. 8th ed. 2009: Wolters Kluwer Health|Lippincott Williams & Wilkins.
16. Chen, P. and J. Liu, *Metabonomics and diabetes mellitus.* Advances in Therapy, 2007. **24**(5): p. 1036-1045.
17. Zhou, J., B. Xu, J. Huang, X. Jia, J. Xue, X. Shi, L. Xiao and W. Li, *1H NMR-based metabonomic and pattern recognition analysis for detection of oral squamous cell carcinoma.* Clinica Chimica Acta, 2009. **401**(1-2): p. 8-13.

18. Bogdanov, M., W.R. Matson, L. Wang, T. Matson, R. Saunders-Pullman, S.S. Bressman and M.F. Beal, *Metabolomic profiling to develop blood biomarkers for Parkinson's disease.* Brain, 2008. **131**: p. 389-396.
19. Beger, R., L. Schnackenberg, R. Holland, D. Li and Y. Dragan, *Metabonomic models of human pancreatic cancer using 1D proton NMR spectra of lipids in plasma.* Metabolomics, 2006. **2**(3): p. 125-134.
20. Lindon, J.C., J.K. Nicholson and E. Holmes, *Handbook of Metabonomics and Metabolomics*. 1st ed. 2007, Amsterdam ; Boston: Elsevier.
21. Vakkilainen, J., G. Steiner, J.-C. Ansquer, F. Aubin, S. Rattier, C. Foucher, A. Hamsten, M.-R. Taskinen and D.G. on behalf of the, *Relationships Between Low-Density Lipoprotein Particle Size, Plasma Lipoproteins, and Progression of Coronary Artery Disease. The Diabetes Atherosclerosis Intervention Study (DAIS).* Circulation, 2003: p. 01.CIR.0000057982.50167.6E.
22. Freedman, D.S., J.D. Otvos, E.J. Jeyarajah, J.J. Barboriak, A.J. Anderson and J.A. Walker, *Relation of Lipoprotein Subclasses as Measured by Proton Nuclear Magnetic Resonance Spectroscopy to Coronary Artery Disease.* Arterioscler Thromb Vasc Biol, 1998. **18**(7): p. 1046-1053.
23. El Harchaoui, K., B.J. Arsenault, R. Franssen, J.P. Despres, K. Hovingh, E.S.G. Stroes, J.D. Otvos, N.J. Wareham, J.J.P. Kastelein, K.T. Khaw and S.M. Boekholdt, *High-Density Lipoprotein Particle Size and Concentration and Coronary Risk.* Annals of Internal Medicine, 2009. **150**(2): p. 84-U43.
24. de Certaines, J.D., L. Le Moyec, F. Seguin, P.A. Eliat and J.M. Constans, *Nuclear magnetic resonance spectroscopy of lipids in cancer.* Current Organic Chemistry, 2007. **11**(6): p. 529-546.
25. Fossel, E.T., J.M. Carr and J. McDonagh, *Detection of Malignant Tumors - Water-Suppressed Proton Nuclear-Magnetic-Resonance Spectroscopy of Plasma.* New England Journal of Medicine, 1986. **315**(22): p. 1369-1376.
26. Otvos, J.D., E.J. Jeyarajah, L.W. Hayes, D.S. Freedman, N.A. Janjan and T. Anderson, *Relationships between the Proton Nuclear Magnetic Resonance Properties of Plasma Lipoproteins and Cancer.* Clinical Chemistry, 1991. **37**(3): p. 369-376.
27. Festa, A., K. Williams, A.J.G. Hanley, J.D. Otvos, D.C. Goff, L.E. Wagenknecht and S.M. Haffner, *Nuclear Magnetic Resonance Lipoprotein Abnormalities in Prediabetic Subjects in the Insulin Resistance Atherosclerosis Study.* Circulation, 2005. **111**(25): p. 3465-3472.
28. Colhoun, H.M., J.D. Otvos, M.B. Rubens, M.R. Taskinen, S.R. Underwood and J.H. Fuller, *Lipoprotein subclasses and particle sizes and their relationship with coronary artery calcification in men and women with and without type 1 diabetes.* Diabetes, 2002. **51**(6): p. 1949-1956.
29. Goff, D.C., R.B. D'Agostino, S.M. Haffner and J.D. Otvos, *Insulin resistance and adiposity influence lipoprotein size and subclass concentrations. Results from the Insulin Resistance Atherosclerosis Study.* Metabolism-Clinical and Experimental, 2005. **54**(2): p. 264-270.
30. Hodge, A.M., A.J. Jenkins, D.R. English, K. O'Dea and G.G. Giles, *NMR-determined lipoprotein subclass profile predicts type 2 diabetes.* Diabetes Research and Clinical Practice, 2009. **83**(1): p. 132-139.

31. Stobo J. D., H.D.B., Ladenson P. W., Traill T. A., Petty B. G., *The principles and practice of medicine* 1996: Appleton & Lange.
32. Huda, S.S., N. Sattar and D.J. Freeman, *Lipoprotein metabolism and vascular complications in pregnancy.* Clinical Lipidology, 2009. **4**(1): p. 91-102.
33. Sanchez-Vera, I., B. Bonet, M. Viana, A. Quintanar, M.D. Martin, P. Blanco, S. Donnay and M. Albi, *Changes in plasma lipids and increased low-density lipoprotein susceptibility to oxidation in pregnancies complicated by gestational diabetes: consequences of obesity.* Metabolism-Clinical and Experimental, 2007. **56**(11): p. 1527-1533.
34. Nelson, D.L. and M.M. Cox, *Lehninger Principles of Biochemistry*. 4th ed. 2005: W H Freeman & Co Publication
35. Rifai, N., G.R. Warnick and M.H. Dominiczak, *Handbook of lipoprotein testing* 2nd ed. 2000: American Association for Clinical Chemistry, Inc.
36. Daykin, C.A., O. Corcoran, S.H. Hansen, I. Bjornsdottir, C. Cornett, S.C. Connor, J.C. Lindon and J.K. Nicholson, *Application of directly coupled HPLC MMR to separation and characterization of lipoproteins from human serum.* Analytical Chemistry, 2001. **73**(6): p. 1084-1090.
37. Alakorpela, M., Y. Hiltunen, J. Jokisaari, S. Eskelinen, K. Kiviniitty, M.J. Savolainen and Y.A. Kesaniemi, *A Comparative Study of 'H NMR Lineshape Fitting Analyses and Biochemical Lipid Analyses of the Lipoprotein Fractions VLDL, LDL and HDL, and Total Human Blood Plasma.* NMR in Biomedicine, 1993. **6**(3): p. 225-233.
38. Alakorpela, M., A. Korhonen, J. Keisala, S. Horkko, P. Korpi, L.P. Ingman, J. Jokisaari, M.J. Savolainen and Y.A. Kesaniemi, *H-1 NMR-Based Absolute Quantitation of Human Lipoproteins and Their Lipid Contents Directly from Plasma.* Journal of Lipid Research, 1994. **35**(12): p. 2292-2304.
39. Ala-Korpela, M., N. Lankinen, A. Salminen, T. Suna, P. Soininen, R. Laatikainen, P. Ingman, M. Jauhiainen, M.R. Taskinen, K. Heberger and K. Kaski, *The inherent accuracy of H-1 NMR spectroscopy to quantify plasma lipoproteins is subclass dependent.* Atherosclerosis, 2007. **190**(2): p. 352-358.
40. Otvos, J., E. Jeyarajah, D. Bennett and R. Krauss, *Development of a proton nuclear magnetic resonance spectroscopic method for determining plasma lipoprotein concentrations and subspecies distributions from a single, rapid measurement.* Clin Chem, 1992. **38**(9): p. 1632-1638.
41. Otvos, J.D., E.J. Jeyarajah and D.W. Bennett, *Quantification of Plasma Lipoproteins by Proton Nuclear Magnetic Resonance Spectroscopy.* Clinical Chemistry, 1991. **37**(3): p. 377-386.
42. Nicholson, J.K., J.C. Lindon and E. Holmes, *'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data.* Xenobiotica, 1999. **29**(11): p. 1181-1189.
43. Bales, J.R., D.P. Higham, I. Howe, J.K. Nicholson and P.J. Sadler, *Use of high resolution proton nuclear magnetic resonance spectroscopy for rapid multi-component analysis of urine.* Clinical Chemistry, 1984. **30**(3): p. 426-432.
44. Nicholson, J.K., M.J. Buckingham and P.J. Sadler, *High-Resolution H-1-Nmr Studies of Vertebrate Blood and Plasma.* Biochemical Journal, 1983. **211**(3): p. 605-615.

45. Nicholson, J.K. and I.D. Wilson, *High resolution proton NMR spectroscopy of biological fluids.* Progress in Nuclear Magnetic Resonance Spectroscopy, 1989. **21**: p. 449-501.
46. Collino, S., F.-P. Martin, S. Kochhar and S. Rezzi, *Monitoring Healthy Metabolic Trajectories with Nutritional Metabonomics.* Nutrients, 2009. **1**(1): p. 101-110.
47. Lindon, J.C., E. Holmes, M.E. Bollard, E.G. Stanley and J.K. Nicholson, *Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis.* Biomarkers, 2004. **9**(1): p. 1-31.
48. Nicholson, J.K., J. Connelly, J.C. Lindon and E. Holmes, *Metabonomics: a platform for studying drug toxicity and gene function.* Nature Reviews Drug Discovery, 2002. **1**(2): p. 153-161.
49. Rousseau, R., B. Govaerts, M. Verleysen and B. Boulanger, *Comparison of some chemometric tools for metabonomics biomarker identification.* Chemometrics and Intelligent Laboratory Systems, 2008. **91**(1): p. 54-66.
50. Sitter, B., U. Sonnewald, M. Spraul, H.E. Fjösne and I.S. Gribbestad, *High-resolution magic angle spinning MRS of breast cancer tissue.* NMR in Biomedicine, 2002. **15**(5): p. 327-337.
51. Sitter, B., T.F. Bathen, T.E. Singstad, H.E. Fjosne, S. Lundgren, J. Halgunset and I.S. Gribbestad, *Quantification of metabolites in breast cancer patients with different clinical prognosis using HR MAS MR spectroscopy.* NMR in Biomedicine, 2010. **23**(4): p. 424-431.
52. Chen, J., W.Z. Wang, S. Lv, P.Y. Yin, X.J. Zhao, X. Lu, F.X. Zhang and G.W. Xu, *Metabonomics study of liver cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations.* Analytica Chimica Acta, 2009. **650**(1): p. 3-9.
53. Yang, Q., X. Shi, Y. Wang, W. Wang, H. He, X. Lu and G. Xu, *Urinary metabonomic study of lung cancer by a fully automatic hyphenated hydrophilic interaction/RPLC-MS system.* Journal of Separation Science, 2010. **33**(10): p. 1495-1503.
54. Rocha, C.u.M., A.n.S. Barros, A.M. Gil, B.J. Goodfellow, E. Humpfer, M. Spraul, I.M. Carreira, J.B. Melo, J.o. Bernardo, A. Gomes, V. Sousa, L. Carvalho and I.F. Duarte, *Metabolic Profiling of Human Lung Cancer Tissue by 1H High Resolution Magic Angle Spinning (HRMAS) NMR Spectroscopy.* Journal of Proteome Research, 2009. **9**(1): p. 319-332.
55. Brindle, J.T., H. Antti, E. Holmes, G. Tranter, J.K. Nicholson, H.W.L. Bethell, S. Clarke, P.M. Schofield, E. McKilligin, D.E. Mosedale and D.J. Grainger, *Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using H-1-NMR-based metabonomics.* Nature Medicine, 2002. **8**(12): p. 1439-1444.
56. Brindle, J.T., J.K. Nicholson, P.M. Schofield, D.J. Grainger and E. Holmes, *Application of chemometrics to H-1 NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension.* Analyst, 2003. **128**(1): p. 32-36.
57. Turner, E., J.A. Brewster, N.A.B. Simpson, J.J. Walker and J. Fisher, *Plasma from women with Preeclampsia has a low lipid and ketone body content - A nuclear magnetic resonance study.* Hypertension in Pregnancy, 2007. **26**(3): p. 329-342.

58. Turner, E., J.A. Brewster, N.A.B. Simpson, J.J. Walker and J. Fisher, *Aromatic amino acid biomarkers of preeclampsia - A nuclear magnetic resonance investigation.* Hypertension in Pregnancy, 2008. **27**(3): p. 225-235.
59. Graca, G., I.F. Duarte, A.S. Barros, B.J. Goodfellow, S. Diaz, I.M. Carreira, A.B. Couceiro, E. Galhano and A.M. Gil, *H-1 NMR Based Metabonomics of Human Amniotic Fluid for the Metabolic Characterization of Fetus Malformations.* Journal of Proteome Research, 2009. **8**(8): p. 4144-4150.
60. Groenen, P.M.W., U.F. Engelke, R.A. Wevers, J.C.M. Hendriks, T.K.A.B. Eskes, H.M.W.M. Merkus and R.P.M. Steegers-Theunissen, *High-resolution 1H NMR spectroscopy of amniotic fluids from spina bifida fetuses and controls.* European Journal of Obstetrics & Gynecology and Reproductive Biology, 2004. **112**(1): p. 16-23.
61. Bock, J., *Metabolic profiling of amniotic fluid by proton nuclear magnetic resonance spectroscopy: correlation with fetal maturation and other clinical variables.* Clin Chem, 1994. **40**(1): p. 56-61.
62. Jacobsen, N.E., *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. 1st ed. 2007, Hoboken, New Jersey: John Wiley & Sons.
63. Graca, G., I.F. Duarte, B.J. Goodfellow, I.M. Carreira, A.B. Couceiro, M.D. Domingues, M. Spraul, L.H. Tseng and A.M. Gil, *Metabolite profiling of human amniotic fluid by hyphenated nuclear magnetic resonance spectroscopy.* Analytical Chemistry, 2008. **80**(15): p. 6085-6092.
64. Duarte, I., A. Barros, P.S. Belton, R. Righelato, M. Spraul, E. Humpfer and A.M. Gil, *High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer.* Journal of Agricultural and Food Chemistry, 2002. **50**(9): p. 2475-2481.
65. Duarte, I.F., I. Delgadillo and A.M. Gil, *Study of natural mango juice spoilage and microbial contamination with Penicillium expansum by high resolution 1H NMR spectroscopy.* Food Chemistry, 2006. **96**(2): p. 313-324.
66. Balci, M., *Basic 1H- and 13C-NMR Spectroscopy*. 2005: Elsevier.
67. Macomber, R.S., *A Complete Introduction to Modern NMR Spectroscopy*. 1998, New York: John Wiley & Sons, Inc.
68. Lambert, J.B. and E.P. Mazzola, *Nuclear Magnetic Resonance Spectroscopy: An Introduction to Principles, Applications and Experimental Methods*. 2003, New Jersey: Pearson Education Inc.
69. Maher, A.D., D. Crockford, H. Toft, D. Malmodin, J.H. Faber, M.I. McCarthy, A. Barrett, M. Allen, M. Walker, E. Holmes, J.C. Lindon and J.K. Nicholson, *Optimization of Human Plasma 1H NMR Spectroscopic Data Processing for High-Throughput Metabolic Phenotyping Studies and Detection of Insulin Resistance Related to Type 2 Diabetes.* Analytical Chemistry, 2008. **80**(19): p. 7354-7362.
70. Nicholson, J.K., P.J.D. Foxall, M. Spraul, R.D. Farrant and J.C. Lindon, *750-MHz 1H AND 1H-13C NMR Spectroscopy of Human Blood Plasma.* Analytical Chemistry, 1995. **67**(5): p. 793-811.

71. Gil, A.M., I. Duarte, E. Cabrita, B.J. Goodfellow, M. Spraul and R. Kerssebaum, *Exploratory applications of diffusion ordered spectroscopy to liquid foods: an aid towards spectral assignment.* Analytica Chimica Acta, 2004. **506**(2): p. 215-223.
72. Pavia, D.L., G.M. Lampman, G.S. Kriz and J.R. Vyvyan, *Introduction to Spectroscopy*. 2009, Brooks/Cole, Cengage Learning: Belmont.
73. Stuart, B.H., *Infrared Spectroscopy: Fundamentals and Applications*. Analytical Techniques in the Sciences (AnTS) 2004: John Wiley & Sons, Ltd.
74. Déléris, G. and C. Petibois, *Applications of FT-IR spectrometry to plasma contents analysis and monitoring.* Vibrational Spectroscopy, 2003. **32**(1): p. 129-136.
75. Erukhimovitch, V., M. Talyshinsky, Y. Souprun and M. Huleihel, *FTIR spectroscopy examination of leukemia patients plasma.* Vibrational Spectroscopy, 2006. **40**(1): p. 40-46.
76. Schmitt, J., M. Beekes, A. Brauer, T. Udelhoven, P. Lasch and D. Naumann, *Identification of Scrapie Infection from Blood Serum by Fourier Transform Infrared Spectroscopy.* Analytical Chemistry, 2002. **74**(15): p. 3865-3868.
77. Lindon, J.C., E. Holmes and J.K. Nicholson, *Metabonomics and its role in drug development and disease diagnosis.* Expert Review of Molecular Diagnostics, 2004. **4**(2): p. 189-199.
78. Fawcett, T., *An introduction to ROC analysis.* Pattern Recogn. Lett., 2006. **27**(8): p. 861-874.
79. Lindon, J.C., J.K. Nicholson, E. Holmes and J.R. Everett, *Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids.* Concepts in Magnetic Resonance, 2000. **12**(5): p. 289-320.
80. Coen, M., E. Holmes, J.C. Lindon and J.K. Nicholson, *NMR-Based Metabolic Profiling and Metabonomic Approaches to Problems in Molecular Toxicology.* Chemical Research in Toxicology, 2008. **21**(1): p. 9-27.
81. Lindon, J.C., E. Holmes and J.K. Nicholson, *Pattern recognition methods and applications in biomedical magnetic resonance.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2001. **39**(1): p. 1-40.
82. Trygg, J., E. Holmes and T. Lundstedt, *Chemometrics in metabonomics.* Journal of Proteome Research, 2007. **6**(2): p. 469-479.
83. Rosipal R., K.N., *Overview and recent advances in partial least squares*, in *Saunders C., Grobelnik M., Gunn S., Shawe-Taylor J., Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop (SLSFS 2005)*. 2006, Springer-Verlag: New York. p. 34-51.
84. Barker, M. and W. Rayens, *Partial least squares for discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166-173.
85. Larsen, F.H., F.v.d. Berg and S.B. Engelsen, *An exploratory chemometric study of 1H NMR spectra of table wines.* Journal of Chemometrics, 2006. **20**(5): p. 198-208.
86. Leardi, R. and L. Nørgaard, *Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions.* Journal of Chemometrics, 2004. **18**(11): p. 486-497.
87. Beckonert, O., H.C. Keun, T.M.D. Ebbels, J. Bundy, E. Holmes, J.C. Lindon and J.K. Nicholson, *Metabolic profiling, metabolomic and metabonomic procedures for*

NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat. Protocols, 2007. **2**(11): p. 2692-2703.

88. Barros, A.S., *Contribution à la sélection et la comparaison de variables caractéristiques.*, in *Institut National Agronomique Paris-Grignon*. 1999: Paris.
89. Harris, R., S.U. Patel, P.J. Sadler and J.H. Viles, *Observation of albumin resonances in proton nuclear magnetic resonance spectra of human blood plasma: N-terminal assignments aided by use of modified recombinant albumin.* Analyst, 1996. **121**(7): p. 913-922.
90. Liu, M.L., J.K. Nicholson, J.A. Parkinson and J.C. Lindon, *Measurement of biomolecular diffusion coefficients in blood plasma using two-dimensional H-1-H-1 diffusion-edited total-correlation NMR spectroscopy.* Analytical Chemistry, 1997. **69**(8): p. 1504-1509.
91. Lan, W., H. Zhu, Z. Zhou, C. Ye and M. Liu, *1H NMR investigation on interaction between ibuprofen and lipoproteins.* Chemistry and Physics of Lipids, 2007. **148**(2): p. 105-111.
92. Murphy, H.C., M. Ala-Korpela, J.J. White, A. Raoof, J.D. Bell, M.L. Barnard, S.P. Burns and R.A. Iles, *Evidence for Distinct Behaviour of Phosphatidylcholine and Sphingomyelin at the Low Density Lipoprotein Surface.* Biochemical and Biophysical Research Communications, 1997. **234**(3): p. 733-737.
93. Otvos, J.D., E.J. Jeyarajah and W.C. Cromwell, *Measurement issues related to lipoprotein heterogeneity.* The American Journal of Cardiology, 2002. **90**(8, Supplement 1): p. 22-29.
94. Zambon, A., P. Simioni, S. Bertocco, M. Marchiori, D. Vianello, D. Tormene, P. Prandoni and A. Pagnan, *Increased levels of small, dense, oxidized LDL in patients with idiopathic venous thromboembolism: A common link with atherosclerosis.* Atherosclerosis Supplements, 2006. **7**(3): p. 502-502.
95. Molloy, A.M., J.L. Mills, C. Cox, S.F. Daly, M. Conley, L.C. Brody, P.N. Kirke, J.M. Scott and P.M. Ueland, *Choline and homocysteine interrelations in umbilical cord and maternal plasma at delivery.* Am J Clin Nutr, 2005. **82**(4): p. 836-842.
96. Wang, C., H. Kong, Y. Guan, J. Yang, J. Gu, S. Yang and G. Xu, *Plasma Phospholipid Metabolic Profiling and Biomarkers of Type 2 Diabetes Mellitus Based on High-Performance Liquid Chromatography/Electrospray Mass Spectrometry and Multivariate Statistical Analysis.* Analytical Chemistry, 2005. **77**(13): p. 4108-4116.
97. Alton, G., M. Hasilik, R. Niehues, K. Panneerselvam, J. Etchison, F. Fana and H. Freeze, *Direct utilization of mannose for mammalian glycoprotein biosynthesis.* Glycobiology, 1998. **8**(3): p. 285-295.
98. Liu, Z.Z., F.A. Carone, T.M. Dalecki, B. Lelongt, E.I. Wallner and Y.S. Kanwar, *Mannose-induced Dysmorphogenesis of Metanephric Kidney - Role of Proteoglycans and Adenosine Triphosphate.* Journal of Clinical Investigation, 1992. **90**(4): p. 1205-1218.
99. von Versen-Hoeynck, F.M. and R.W. Powers, *Maternal-fetal metabolism in normal pregnancy and preeclampsia.* Frontiers in Bioscience, 2007. **12**: p. 2457-2470.